



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

<p>(51) Classification internationale des brevets ⁶ :</p> <p>C12N 15/74, 15/31, C07K 14/35, A61K 48/00, 39/04, C07K 19/00, C12Q 1/68, C07K 16/12, G01N 33/50, 33/53 // C12N 15/52, 15/65</p>	<p>A2</p>	<p>(11) Numéro de publication internationale: WO 99/09186</p> <p>(43) Date de publication internationale: 25 février 1999 (25.02.99)</p>						
<p>(21) Numéro de la demande internationale: PCT/FR98/01813</p> <p>(22) Date de dépôt international: 14 août 1998 (14.08.98)</p> <p>(30) Données relatives à la priorité:</p> <table border="0"> <tr> <td>97/10404</td> <td>14 août 1997 (14.08.97)</td> <td>FR</td> </tr> <tr> <td>97/11325</td> <td>11 septembre 1997 (11.09.97)</td> <td>FR</td> </tr> </table> <p>(71) Déposant (pour tous les Etats désignés sauf US): INSTITUT PASTEUR [FR/FR]; 28, rue du Docteur Roux, F-75015 Paris (FR).</p> <p>(72) Inventeurs; et</p> <p>(75) Inventeurs/Déposants (US seulement): GICQUEL, Brigitte [FR/FR]; 8, rue Daguerre, F-75014 Paris (FR). PORTNOI, Denis [FR/FR]; 7, rue Simon Lefranc, F-75004 Paris (FR). LIM, Eng-Mong [KH/FR]; 20, rue Georges Pitard, F-75015 Paris (FR). PELJICIC, Vladimir [FR/FR]; 28, rue de Chateaudun, F-75009 Paris (FR). GUIGUENO, Agnès [FR/FR]; 26-28, rue Gambetta, F-62026 Arras (FR). GOGUET DE LA SALMONIERE, Yves [FR/FR]; 30, rue Lourmel, F-75015 Paris (FR).</p>		97/10404	14 août 1997 (14.08.97)	FR	97/11325	11 septembre 1997 (11.09.97)	FR	<p>(74) Mandataires: MARTIN, Jean-Jacques etc.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).</p> <p>(81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Publiée</p> <p><i>Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.</i></p> <p><i>Avec une indication relative à du matériel biologique déposé, fournie selon la règle 13bis, séparément, et non avec la description.</i></p>
97/10404	14 août 1997 (14.08.97)	FR						
97/11325	11 septembre 1997 (11.09.97)	FR						
<p>(54) Title: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA, VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND PREVENTING TUBERCULOSIS</p>								
<p>(54) Titre: SEQUENCES NUCLEIQUES DE POLYPEPTIDES EXPORTES DE MYCOBACTERIES, VECTEURS LES COMPRENANT ET APPLICATIONS AU DIAGNOSTIC ET A LA PREVENTION DE LA TUBERCULOSE</p>								
<p>(57) Abstract</p> <p>The invention concerns recombinant vectors replicated in mycobacteria, a set of sequences coding for exported polypeptides detected by fusion with alkaline phosphatase, in particular one polypeptide, called DP428, of about 12 kD corresponding to an exported protein found in mycobacteria belonging to the <i>Mycobacterium tuberculosis</i> complex. The invention also concerns methods and kits for detecting in vitro the presence of a mycobacterium and in particular a mycobacterium belonging to the <i>Mycobacterium tuberculosis</i> complex in a biological sample using said polypeptides, their fragments or polynucleotides coding for the latter. The invention also concerns immunogenic or vaccine compositions for preventing and/or treating infections caused by mycobacteria and in particular a mycobacterium belonging to said complex, particularly tuberculosis.</p>								
<p>(57) Abrégé</p> <p>L'invention a pour objet des vecteurs recombinants se répliquant chez les mycobactéries, un ensemble de séquences codant pour des polypeptides exportés détectés par des fusions avec la phosphatase alcaline, notamment un polypeptide, dénommé DP428, d'environ 12kD correspondant à une protéine exportée retrouvée dans les mycobactéries appartenant au complexe de <i>Mycobacterium tuberculosis</i>. L'invention concerne également des procédés et des kits de détection <i>in vitro</i> de la présence d'une mycobactérie et en particulier une mycobactérie appartenant au complexe de <i>Mycobacterium tuberculosis</i> dans un échantillon biologique utilisant lesdits polypeptides, leurs fragments ou des polynucléotides codant pour ces derniers. L'invention vise des compositions immunogènes ou vaccins pour la prévention et/ou le traitement d'infections provoquées par des mycobactéries et en particulier une mycobactérie appartenant audit complexe, en particulier la tuberculose.</p>								

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brézil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Biélorus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroon	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

Séquences nucléiques de polypeptides exportés de mycobactéries, vecteurs les comprenant et applications au diagnostic et à la prévention de la tuberculose.

5 L'invention a pour objet de nouveaux vecteurs recombinants de criblage, de clonage et/ou d'expression se répliquant chez les mycobactéries. Elle a également pour objet un ensemble de séquences codant pour des polypeptides exportés détectés par des fusions avec la phosphatase
10 alcaline et dont l'expression est régulée (induite ou réprimée) ou constitutive lors de l'ingestion des mycobactéries par les macrophages. L'invention concerne également un polypeptide, dénommé DP428, d'environ 12kD correspondant à une protéine exportée retrouvée dans les
15 mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*. L'invention vise aussi un polynucléotide comprenant une séquence codant pour ce polypeptide. Elle concerne également l'utilisation du polypeptide ou de fragments de celui-ci et des polynucléotides codant pour
20 ces derniers (ou encore les polynucléotides complémentaires à ces derniers) pour la réalisation de moyens de détection in vitro, ou in vivo de la présence d'une mycobactérie appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique ou pour la détection de réactions
25 de l'hôte infecté par ces espèces bactériennes. L'invention vise enfin l'utilisation du polypeptide ou de fragments de celui-ci ainsi que des polynucléotides codant pour ces derniers en tant que moyens destinés à la préparation d'une composition immunogène, susceptible d'induire une réponse
30 immunitaire dirigée contre les mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*, ou d'une composition vaccinale pour la prévention et/ou le traitement d'infections provoquées par des mycobactéries appartenant audit complexe, en particulier la tuberculose.

La présente invention a aussi pour but d'utiliser ces séquences (polypeptidiques et polynucléotidiques) comme
5 cible pour la recherche de nouveaux inhibiteurs de la croissance et de la multiplication des mycobactéries et de leur maintien chez l'hôte, ses inhibiteurs pouvant servir d'antibiotiques.

10 Le genre *Mycobacterium*, qui comprend au moins 56 espèces différentes, inclut des pathogènes humains majeurs tels que *M. leprae* et *M. tuberculosis*, les agents responsables de la lèpre et de la tuberculose, qui restent des problèmes graves de santé publique dans le monde
15 entier.

La tuberculose continue d'être un problème de santé publique dans le monde. Aujourd'hui, cette maladie est la cause de 2 à 3 millions de morts dans le monde et environ 8
20 millions de nouveaux cas sont observés chaque année (Bouvet, 1994). Dans les pays développés *M. tuberculosis* est la cause la plus commune des infections mycobactériennes. En France il apparaît environ 10 000 nouveaux cas par an et parmi les maladies à déclaration
25 obligatoire c'est la tuberculose qui comprend le plus grand nombre de cas. La vaccination par le BCG (Bacille de Calmette et Guérin), une souche avirulente dérivée de *M. bovis* et qui est très utilisée comme vaccin contre la tuberculose, est loin d'être efficace au sein de toutes les
30 populations. Cette efficacité varie environ de 80 % dans les pays occidentaux comme l'Angleterre, à 0 % en Inde (résultats du dernier essai de vaccination à Chingleput., publiés en 1972 dans *Indian J. Med. Res.*). De plus, l'apparition de souches de *M. tuberculosis* résistantes aux
35 antituberculeux et le risque accru chez les patients immunodéprimés, patients atteints du SIDA, de développer

une tuberculose, rendent nécessaire la mise au point de méthodes rapides, spécifiques et fiables pour le diagnostic de la tuberculose et la mise au point de nouveaux vaccins. Par exemple, une étude épidémiologique réalisée en Floride, et dont les résultats ont été publiés en 1993 dans AIDS thérapies, a montré que 10 % des malades atteints de SIDA sont atteints de tuberculose au moment du diagnostic du SIDA ou 18 mois avant celui-ci. Chez ces malades, la tuberculose apparaît dans 60 % des cas sous une forme disséminée donc non repérable par les critères de diagnostic classiques comme la radiographie pulmonaire ou l'analyse de crachats.

Actuellement, une certitude sur le diagnostic apporté par la mise en évidence de bacilles cultivables dans un prélèvement provenant du malade n'est obtenue que pour moins de la moitié des cas de tuberculose, même dans les cas de tuberculose pulmonaire. Le diagnostic de la tuberculose et des autres mycobactéries apparentées est donc difficile à réaliser, et cela pour différentes raisons : les mycobactéries sont souvent présentes en faible quantité, leur temps de génération est très long (24h pour *M. tuberculosis*) et leur culture est difficile. (Bates et al., 1986).

D'autres techniques sont utilisables en clinique, pour identifier une infection mycobactérienne :

a) L'identification directe des microorganismes au microscope ; cette technique est rapide, mais ne permet pas l'identification de l'espèce mycobactérienne observée et manque de sensibilité (Bates, 1979).

Les cultures, lorsqu'elles sont positives, ont une spécificité approchant 100 % et permettent l'identification de l'espèce mycobactérienne isolée ; néanmoins, comme précisé ci-dessus, la croissance des mycobactéries *in vitro* est longue (ne peut être réalisée qu'en 3 à 6 semaines de

cultures répétées (Bates, 1979 ; Bates et al., 1986)) et coûteuse.

b) Les techniques sérologiques peuvent s'avérer utiles dans certaines conditions, mais leur utilisation est parfois limitée par leur sensibilité et/ou leur spécificité faibles (Daniel et al., 1987).

c) La présence de mycobactéries au sein d'un échantillon biologique peut aussi être déterminée par hybridation moléculaire avec de l'ADN ou de l'ARN en utilisant des sondes d'oligonucléotides spécifiques des séquences recherchées (Kiehn et al., 1987 ; Roberts et al., 1987 ; Drake et al., 1987). Plusieurs études ont montré l'intérêt de cette technique pour le diagnostic des infections à mycobactéries. Les sondes utilisées sont constituées d'ADN, d'ARN ribosomique ou de fragments d'ADN mycobactériens provenant de banque de gènes. Le principe de ces techniques repose sur le polymorphisme des séquences nucléotidiques des fragments utilisés ou sur le polymorphisme des régions avoisinantes. Dans tous les cas, elles nécessitent l'utilisation de cultures et ne sont pas applicables directement sur les échantillons biologiques.

La faible quantité de mycobactéries présentes au sein d'un échantillon biologique et en conséquence la quantité faible d'ADN cible à détecter dans cet échantillon peut nécessiter le recours à une amplification spécifique *in vitro* de l'ADN cible avant sa détection à l'aide de la sonde nucléotidique et en utilisant des techniques d'amplification *in vitro* telles que la PCR (amplification en chaîne à la polymérase. L'amplification spécifique de l'ADN par la technique PCR peut constituer la première étape d'un procédé de détection de la présence d'un ADN mycobactérien dans un échantillon biologique, la détection proprement dite de l'ADN amplifié étant effectuée dans un

second temps à l'aide d'une sonde oligonucléotidique capable de s'hybrider spécifiquement à l'ADN amplifié.

Un test de détection de mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*, par hybridation sandwich (test utilisant une sonde de capture et une sonde de détection) a été décrit par Chevrier et al. en 1993. le complexe de *Mycobacterium tuberculosis* est un groupe de mycobactéries qui comprend *M. bovis-BCG*, *M. bovis*, *M. tuberculosis*, *M. africanum* et *M. microti*.

10

Un procédé de détection de faibles quantités de mycobactéries, appartenant au complexe tuberculosis, par amplification génique et hybridation directement sur des échantillons biologiques a été mis au point. Ledit procédé utilise la séquence d'insertion IS6110 (Brevet européen EP 0 490 951 B1). Thierry et al. ont décrit en 1990 une séquence spécifique du complexe *Mycobacterium tuberculosis* et nommée IS 6110. Certains auteurs ont proposé d'amplifier spécifiquement l'ADN provenant de *Mycobacterium* en utilisant des amorces nucléiques dans une méthode d'amplification, telle que la réaction de polymérase en chaîne (PCR). Patel et al. ont décrit en 1990 l'utilisation de plusieurs amorces nucléiques choisies à partir d'une séquence connue en tant que sonde dans l'identification de *M. tuberculosis*. Cependant, la longueur des fragments obtenus en utilisant ces amorces était différente de la longueur théorique attendue et plusieurs fragments de taille variable étaient obtenus. De plus, les auteurs ont observé l'absence d'hybridation des produits amplifiés avec le plasmide ayant servi à déterminer les amorces. Ces résultats indiquent que ces amorces ne seraient pas appropriées dans la détection de la présence de *M. tuberculosis* dans un échantillon biologique et confirment la nature critique du choix des amorces. La même année, J.L. Guesdon et D. Thierry ont décrit une méthode de

35

détection de *M. tuberculosis*, de grande sensibilité, par amplification d'un fragment d'ADN de *M. tuberculosis* localisé au sein de la séquence IS6110 (Brevet européen EP 461 045) à l'aide d'amorces générant des fragments d'ADN amplifié de longueur constante, même lorsque le choix des amorces conduisait à l'amplification de fragments longs (de l'ordre de 1000 à 1500 bases) où le risque d'interruption de la polymérisation est élevée en raison des effets de la structure secondaire de la séquence. D'autres amorces spécifiques de la séquence IS6110 sont décrites dans le brevet européen N° EP-0490 951.

Les inventeurs ont montré (résultats non publiés) que certains isolats cliniques de *Mycobacterium tuberculosis* étaient exempts de la séquence d'insertion IS6110 et ne pouvaient donc être détectés à l'aide des oligonucléotides spécifiques de cette séquence pouvant conduire ainsi à des résultats de diagnostic faussement négatifs. Ces résultats confirment une observation similaire faite par Yuen et al. en 1993. L'impossibilité de détecter ces souches pathogènes potentiellement présentes dans un échantillon biologique prélevé sur un patient est ainsi susceptible de conduire à des difficultés voire des erreurs de diagnostic. La disponibilité de plusieurs séquences spécifiques du Bacille de la tuberculose, à l'intérieur desquelles des amorces appropriées pour l'amplification seront choisis, est importante. La séquence DP428 décrite ici pourra être utilisée.

M. bovis et *M. tuberculosis*, les agents causals de la tuberculose, sont des bactéries facultatives intracellulaires.

Ces agents ont développé des mécanismes pour assurer leur survie et leur répllication à l'intérieur du macrophage, un des types cellulaires qui est supposé

éradiquer l'invasion par des microorganismes. Ces agents sont capables de moduler l'évolution normale de leur phagosome et de les empêcher de se différencier en un compartiment acide riche en hydrolase (Clemens, 1979 ;
5 Clemens et al., 1996; Sturgill-Koszycki et al., 1994 et Xu et al., 1994). Cependant, cette modulation n'est possible que si la bactérie est vivante au sein du phagosome, suggérant que des composés synthétisés de manière active et/ou sécrétés à l'intérieur de la cellule font partie de
10 ce mécanisme. Des protéines exportées sont probablement impliquées dans ce mécanisme. En dépit des problèmes majeurs de santé liés à ces organismes pathogènes, on sait peu de choses sur leurs protéines exportées et/ou sécrétées. Des analyses en SDS-PAGE de filtrat de culture
15 de *M. tuberculosis* montrent au moins 30 protéines sécrétées (Altschul et al., 1990 ; Nagal et al., 1991 et Young et al., 1992). Certaines d'entre elles ont été caractérisées, leurs gènes clonés et séquencés (Borremans et al., 1989 ; Wiker et al., 1992 et Yamaguchi et al., 1989). D'autres,
20 bien qu'il s'agisse d'antigènes immunodominants d'importance majeure pour induire une immunité protectrice (Anderson et al., 1991 et Orme et al., 1993), ne sont pas totalement identifiés. En outre, il est probable que de nombreuses protéines exportées restent fixées sur la
25 membrane cellulaire et par conséquent ne soient pas présentes dans les surnageants de culture. Il a été montré que les protéines localisées à la surface externe de diverses bactéries pathogènes, telles que l'invasine de 103 kDa de *Yersina Pseudotuberculosis* (Isberg et al., 1987) ou
30 l'internaline de 80 kDa de *Listeria monocytogenes* (Gaillard et al., 1991 et Dramsi et al., 1997) jouent un rôle important dans les interactions avec les cellules hôtes et par conséquent, dans la pathogénicité comme dans l'induction de réponses protectrices. Ainsi, une protéine
35 liée à la membrane pourrait être importante pour l'infection à *M. tuberculosis* comme pour l'induction de

réponse protectrice contre cette infection. Ces protéines pourraient revêtir un intérêt certain pour la préparation de vaccins.

5 Récemment, il a été décrit l'adaptation aux mycobactéries d'une méthodologie génétique pour l'identification et la sélection phénotypique de protéines exportées (Lim et al., 1995). Cette méthode utilise la phosphatase alcaline (PhoA) périplasmique d'*E. coli*. Un
10 vecteur plasmidique a été construit permettant la fusion de gènes entre un gène *PhoA* tronqué et des gènes codant pour des protéines exportées (Manoil et al., 1990).

 Par cette méthode, il a pu être identifié un gène
15 de *M. tuberculosis* (*erp* (Berthet et al., 1995)) présentant des homologies avec une protéine exportée de 28 kDa de *M. leprae*, qui est une cible fréquente des réponses humores de la forme lépromateuse de la lèpre. Une protéine présentant des motifs aminoacides caractéristiques de la
20 désaturase de plante (*des*) a aussi été caractérisée par la technique de fusion avec *PhoA*.

 Cependant, cette méthode génétique d'identification de protéines exportées ne permet pas d'évaluer facilement
25 l'expression intracellulaire des gènes correspondants. Une telle évaluation est d'une importance primordiale à la fois pour la sélection de bons candidats vaccins et pour la compréhension des interactions entre les bactéries et leurs cellules hôtes. L'induction de l'expression de facteur de
30 virulence par contact de cellule cible pathogène a été décrite. C'est le cas par exemple pour les facteurs de virulence Yops (Petersson et al., 1996) de *Yersinia pseudotuberculosis*. *Shigella* par contact avec les cellules
35 cibles relargue les protéines Ipa dans le milieu de culture, et *Salmonella* synthétise de nouvelles structures

de surface.

Compte tenu de ce qui précède, il existe aujourd'hui un grand besoin de développer de nouveaux vaccins contre les mycobactéries pathogènes ainsi que de nouveaux tests de diagnostic spécifiques, fiables et rapides. Ces développements nécessitent la mise au point d'outils spécifiques encore plus performants permettant, d'une part, d'isoler ou d'obtenir des séquences de nouveaux polypeptides spécifiques, notamment immunogènes, et, d'autre part, de mieux comprendre le mécanisme des interactions entre les bactéries et leurs cellules hôtes comme notamment l'induction de l'expression de facteur de virulence. Ceci est précisément l'objet de la présente invention.

Les inventeurs ont défini et réalisé dans ce but de nouveaux vecteurs permettant le criblage, le clonage et/ou l'expression de séquences d'ADN de mycobactéries afin d'identifier parmi ces séquences, des acides nucléiques codant pour des protéines d'intérêt, de préférence des protéines exportées, pouvant être localisées sur la membrane bactérienne et/ou sécrétées, et d'identifier parmi ces séquences celles qui sont induites ou réprimées lors de l'infection (croissance intracellulaire).

Description

La présente invention décrit l'utilisation du gène rapporteur *phoA* chez les mycobactéries. Il permet d'identifier des systèmes d'expression et d'exportation dans un contexte mycobactérien. Beaucoup de gènes ne sont exprimés que dans un tel contexte, ce qui montre l'avantage de la présente invention. Au cours du clonage de segments d'ADN de souches du complexe *M. Tuberculosis* en fusion avec *phoA* dans une autre mycobactérie comme *M. smegmatis*, le

début du gène, ses régions régulatrices et son régulateur seront clonés ce qui permettra d'observer une régulation. Si cette régulation est positive, le clonage du régulateur constituera un avantage pour observer l'expression et l'exportation.

Dans le contexte de l'invention, on entend par mycobactérie toutes les mycobactéries appartenant aux diverses espèces énumérées par Wayne L. G. and Kubica G. P. (1980). Family Mycobacteriaceae in Bergey's manual of systematic bacteriology, J. P. Butler Ed. (Baltimore USA : Williams et Wilkins P. 1436-1457).

Dans certains cas les gènes clonés sont soumis dans leur hôte d'origine à une régulation négative rendant l'observation de l'expression et de l'exportation difficile chez l'hôte d'origine. Dans ce cas, le clonage du gène en absence de son régulateur négatif, dans un hôte ne le contenant pas, constituera un avantage.

L'invention vise aussi de nouveaux polypeptides et de nouveaux polynucléotides de mycobactéries ayant pu être isolés au moyen des vecteurs précédents et susceptibles d'entrer dans la réalisation de compositions pour la détection d'une infection par des mycobactéries, ou pour la protection contre une infection due à des mycobactéries ou pour la recherche d'inhibiteurs comme cela est décrit précédemment pour DP428.

L'invention a donc pour objet un vecteur recombinant de criblage, de clonage et/ou d'expression, caractérisé en ce qu'il se réplique chez des mycobactéries et en ce qu'il contient :

- 1) un réplicon fonctionnel chez les mycobactéries ;
- 2) un marqueur de sélection ;

3) une cassette rapporteur comprenant :

- a) un site de clonage multiple (polylinker),
- b) éventuellement un terminateur de transcription
5 actif chez les mycobactéries, en amont du polylinker,
- c) une séquence nucléotidique codante issue d'un gène
codant pour un marqueur d'expression, d'exportation et/ou
de sécrétion de protéine, ladite séquence nucléotidique
étant dépourvue de son codon d'initiation et de ses
10 séquences de régulation, et
- d) une séquence nucléotidique codante issue d'un gène
codant pour un marqueur d'activité de promoteurs contenus
dans le même fragment, ladite séquence nucléotidique étant
pourvue de son codon d'initiation. Eventuellement, le
15 vecteur recombinant contient également un réplicon
fonctionnel chez E. coli.

De manière préférée, le marqueur d'exportation et/ou
de sécrétion est placé dans la même orientation que le
20 marqueur d'activité de promoteurs.

Préférentiellement, le vecteur recombinant de criblage
selon l'invention comprendra, en outre, un terminateur de
transcription placé en aval du marqueur d'activité de
25 promoteurs, ce qui est de nature à permettre l'obtention de
transcrits courts qui se révèlent plus stables et qui, par
conséquent, permettent un plus haut niveau d'expression des
produits de traduction.

30 Le marqueur d'exportation et/ou de sécrétion est une
séquence de nucléotides dont l'expression suivie de
l'exportation et/ou de la sécrétion dépend des éléments de
régulation qui contrôlent son expression.

35 Par "séquences ou éléments de régulation de
l'expression de la production de polypeptides et de sa
localisation", on entend une séquence promotrice de la

transcription, une séquence comprenant le site de liaison au ribosome (RBS), les séquences responsables de l'exportation et/ou la sécrétion telles que la séquence dite séquence signal.

5

Un premier marqueur intéressant d'exportation et/ou d'expression est une séquence codante issue du gène *phoA*. Le cas échéant, elle est tronquée de telle façon que l'activité phosphatase alcaline est cependant susceptible d'être restaurée lorsque la séquence codante tronquée est placée sous le contrôle d'un promoteur et d'éléments de régulation appropriés.

10

D'autres marqueurs d'exposition, d'exportation et/ou de sécrétion peuvent être utilisés. On citera à titre d'exemples une séquence du gène β -agarase, de la nucléase d'un staphylocoque ou d'une β -lactamase.

15

Parmi les marqueurs intéressants d'activité de promoteurs contenus dans le même fragment, on préfère une séquence codante issue du gène *luc* de luciférase de luciole pourvue de son codon d'initiation.

20

D'autres marqueurs d'activité de promoteurs contenus dans le même fragment peuvent être utilisés. On citera à titre d'exemples une séquence du gène de la GFP (Green Fluorescent Protein).

25

Le terminateur de transcription doit être fonctionnel chez les mycobactéries. Un terminateur avantageux est à cet égard le terminateur du coliphage T4 (tT4). D'autres terminateurs appropriés pour la réalisation de l'invention peuvent être isolés en utilisant la technique présentée dans les exemples, par exemple au moyen d'une cassette "omega" (Prentki et al., 1984).

30

35

Un vecteur particulièrement préféré pour la réalisation de l'invention est un plasmide choisi parmi les plasmides suivants déposés à la CNCM (Collection Nationale de Cultures de Microorganismes, 25 rue de Docteur Roux, 5 75724 Paris cedex 15, France) :

a) pJVEDa déposé à la CNCM sous le N° I-1797, le 12/12 1996,

b) pJVEDb déposé à la CNCM sous le N° I-1906, le 25 10 juillet 1997,

c) pJVEDc déposé à la CNCM sous le N° I-1799 , le 12/12 1996.

Pour la sélection, ou l'identification de séquences 15 d'acides nucléiques de mycobactéries codant pour des polypeptides susceptibles d'être incorporés dans des compositions immunogènes, ou antigéniques pour la détection d'une infection, ou susceptibles d'induire ou de réprimer un facteur de virulence de mycobactéries, le vecteur de 20 l'invention comprendra, en l'un des sites de clonage multiple du polylinker, une séquence de nucléotides d'une mycobactérie chez laquelle on détecte la présence de séquences correspondant à des polypeptides exportés et/ou sécrétés pouvant être induits ou réprimés lors de 25 l'infection, ou encore exprimés ou produits de façon constitutive, leurs séquences promotrices et/ou régulatrices associées susceptibles de permettre ou de favoriser l'exportation et/ou la sécrétion desdits polypeptides d'intérêt, ou tout ou partie de gènes 30 d'intérêt codant pour lesdits polypeptides.

De préférence, cette séquence est obtenue par fragmentation physique ou par digestion enzymatique de 1'ADN génomique ou de 1'ADN complémentaire d'un ARN d'une 35 mycobactérie, de préférence *M. tuberculosis* ou choisie parmi *M. africanum*, *M. bovis*, *M. avium* ou *M. leprae*.

Les vecteurs de l'invention peuvent en effet également être utilisés pour déterminer la présence de séquences d'intérêt, de préférence correspondant à des protéines exportées et/ou sécrétées, et/ou capables d'être induites ou réprimées ou produites de façon constitutive lors de l'infection, notamment lors de la phagocytose par les macrophages, et selon ce qui a été exposé précédemment, chez des mycobactéries telles que *M. africanum*, *M. bovis*, *M. avium* ou *M. leprae* dont on aura traité l'ADN ou l'ADNc par fragmentation physique ou avec des enzymes déterminées.

Selon un premier mode de réalisation de l'invention la digestion enzymatique de l'ADN génomique ou de l'ADN complémentaire est effectuée à partir de *M. tuberculosis*.

De préférence cet ADN est digéré avec une enzyme telle que sau3A, BclI, BglII.

D'autres enzymes de digestion telles que ScaI, ApaI, SacII, KpnI ou encore des nucléases ou des polymérases, peuvent naturellement être mises en oeuvre, dès lors qu'elles permettent l'obtention de fragments dont les extrémités peuvent être insérées dans l'un des sites de clonage du polylinker du vecteur de l'invention.

Le cas échéant, des digestions avec différentes enzymes seront effectuées simultanément.

Des vecteurs recombinants préférés pour la réalisation de l'invention sont choisis parmi les vecteurs recombinants suivants déposés à la CNCM :

- a) p6D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1814,
- b) p5A3 déposé le 28 janvier 1997 à la CNCM sous le N°I-1815,
- c) p5F6 déposé le 28 janvier 1997 à la CNCM sous le N°I-1816,
- d) p2A29 déposé le 28 janvier 1997 à la CNCM sous le N°I-1817,

- e) pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818,
f) p5B5 déposé le 28 janvier 1997 à la CNCM sous le N°I-1819,
5 g) p1C7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1820,
h) p2D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1821,
10 i) p1B7 déposé le 31 janvier 1997 à la CNCM sous le N°I-1843,
j) pJVED/M. tuberculosis déposé le 25 juillet 1997 à la CNCM sous le N° I-1907,
k) pM1C25 déposé le 4 août 1998 à la CNCM sous le n°I-2062.

15 Parmi les plus préférés, on préfère le vecteur recombinant pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818, et le vecteur pM1C25 déposé le 4 août 1998 à la CNCM sous le N° I-2062.

20 L'invention a également pour objet un procédé de criblage de séquences de nucléotides issues de mycobactéries pour déterminer la présence de séquences correspondant à des polypeptides exportés et/ou sécrétés
25 pouvant être induits ou réprimés lors de l'infection, leurs séquences promotrices et/ou régulatrices associées susceptibles notamment de permettre ou de favoriser l'exportation et/ou la sécrétion desdits polypeptides d'intérêt, ou tout ou partie de gènes d'intérêt codant pour
30 lesdits polypeptides, caractérisé en ce qu'il met en oeuvre un vecteur recombinant selon l'invention.

L'invention concerne aussi un procédé de criblage, selon l'invention, caractérisé en ce qu'il comprend les étapes suivantes :

- 35 a) la fragmentation physique des séquences d'ADN de mycobactéries ou leur digestion par au moins une enzyme

déterminée et la récupération des fragments obtenus ;

b) l'insertion des fragments obtenus à l'étape a) dans un site de clonage, compatible le cas échéant avec l'enzyme de l'étape a), du polylinker d'un vecteur selon l'invention ;

c) si besoin, l'amplification desdits fragments contenus dans le vecteur, par exemple par réplication de ce dernier après insertion du vecteur ainsi modifié dans une cellule déterminée, de préférence *E coli* ;

d) la transformation des cellules hôtes par le vecteur amplifié à l'étape c), ou en l'absence d'amplification, par le vecteur de l'étape b) ;

e) la culture des cellules hôtes transformées dans un milieu permettant la mise en évidence du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs contenu dans le vecteur ;

f) la détection des cellules hôtes positives (colonies positives) pour l'expression du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs ;

g) l'isolement de l'ADN des colonies positives et l'insertion de cet ADN dans une cellule identique à celle de l'étape c) ;

h) la sélection des insertions contenues dans le vecteur, permettant l'obtention de clones positifs pour le marqueur d'exportation et/ou de sécrétion, et /ou pour le marqueur d'activité de promoteurs ;

i) l'isolement et la caractérisation des fragments d'ADN de mycobactéries contenues dans ces insérats.

30

Dans l'un des modes de réalisation préférés du procédé de criblage selon l'invention, les cellules hôtes positives, détectées à l'étape f), pour le marqueur d'exportation et/ou de sécrétion sont, éventuellement dans un second temps, testées pour la capacité de l'insert nucléotidique sélectionné à stimuler l'expression du

35

marqueur d'activité de promoteurs lorsque lesdites cellules hôtes sont phagocytées par des cellules du type macrophagique.

5 De manière plus spécifique, on compare la stimulation de l'expression du marqueur d'activité de promoteurs chez des cellules hôtes placées en culture axénique (cellules hôtes seules en culture) à la stimulation de l'expression du marqueur d'activité de promoteurs chez des cellules
10 hôtes cultivées en présence de macrophages et ainsi phagocytées par ces derniers.

La sélection de cellules hôtes positives pour le marqueur d'activité de promoteurs peut être réalisée dès
15 l'étape e) du procédé de criblage décrit ci-dessus, ou encore après l'une quelconque des étapes f), g), h) ou i), c'est-à-dire une fois que les cellules hôtes ont été sélectionnées positivement pour le marqueur d'exportation et/ou de sélection.

20 La mise en oeuvre de ce procédé permet la construction de banques d'ADN comportant des séquences correspondant à des polypeptides susceptibles d'être exportés et/ou sécrétés, et/ou susceptibles d'être induits ou réprimés
25 lors de l'infection lorsqu'ils sont produits au sein de mycobactéries recombinantes. L'étape i) du procédé peut comprendre une étape de séquençage des insertions sélectionnées.

30 De préférence, dans le procédé selon l'invention, le vecteur utilisé est choisi parmi les plasmides pJVEDa (CNCM, N° I-1797), pJVEDb (CNCM, N° I-1906), pJVEDc (CNCM, N° I-1799) ou pJVED/M. tuberculosis (CNCM, N° I-1907), et la digestion des séquences d'ADN de mycobactéries est
35 effectuée au moyen de l'enzyme Sau3A.

Selon un mode de réalisation préféré de l'invention,

le procédé de criblage est caractérisé en ce que les séquences de mycobactéries sont issues d'une mycobactérie pathogène, par exemple de *M. tuberculosis*, *M. bovis*, *M. avium*, *M. africanum* ou *M. leprae*.

5

L'invention comprend également une banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie, caractérisée en ce qu'elle est obtenue par un procédé comprenant les étapes a) et b), ou a), b) et c) du procédé
10 précédent selon l'invention, de préférence une banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactéries pathogènes, de préférence de mycobactéries appartenant au groupe du complexe *Mycobacterium tuberculosis*, de préférence de *Mycobacterium tuberculosis*.

15

Dans la présente invention, on entend désigner par "séquences nucléiques" ou "séquences d'acides aminés" SEQ ID N° X à SEQ ID N° Y, où X et Y peuvent représenter indépendamment un nombre ou un caractère alphanumérique,
20 respectivement l'ensemble des séquences nucléiques ou l'ensemble des séquences d'acides aminés représentées par les figures X à Y, extrémités comprises.

Par exemple, les séquences nucléiques ou les séquences
25 d'acides aminés SEQ ID N° 1 à SEQ ID N° 4N sont respectivement les séquences nucléiques ou les séquences d'acides aminés représentées par les figures 1 à 4N, c'est-à-dire respectivement les séquences nucléiques ou les séquences d'acides aminés SEQ ID N° 1, SEQ ID N° 1A', SEQ
30 ID N° 1B', SEQ ID N° 1C', SEQ ID N° 1D, SEQ ID N° 1F, SEQ ID N° 2, SEQ ID N° 3A, SEQ ID N° 3B, SEQ ID N° 3C, SEQ ID N° 4A, SEQ ID N° 4B, SEQ ID N° 4C, SEQ ID N° 4A', SEQ ID N° 4B', SEQ ID N° 4C', SEQ ID N° 4F, SEQ ID N° 4J, SEQ ID N° 4K, SEQ ID N° 4L, SEQ ID N° 4M et SEQ ID N° 4N.

35

L'invention a également pour objet les séquences

nucléotidiques de mycobactéries ou comprenant des séquences nucléotidiques de mycobactéries sélectionnées après la réalisation du procédé selon l'invention ci-dessus décrit.

5

De préférence, ladite mycobactérie est choisie parmi *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. avium*, *M. leprae*, *M. paratuberculosis*, *M. kansasii* ou *M. xenopi*.

10

On préfère les séquences nucléotidiques de mycobactéries ou comprenant une séquence nucléotidique de mycobactérie, ladite séquence nucléotidique de mycobactérie étant choisie parmi les séquences de fragments d'ADN de mycobactérie de séquence nucléique SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F, respectivement représentées par les figures 1 à 24C (planches 1 à 150), par les figures 27A à 27C (planches 152 à 154), par la figure 29 (planche 156) et par les figures 31A à 50F (planches 158 à 275).

20

Selon un mode de réalisation particulier de l'invention, des séquences préférées sont par exemple les fragments d'ADN de mycobactéries de séquence SEQ ID N°1, SEQ ID N°3A, SEQ ID N°5A, SEQ ID N°6A, SEQ ID N°7A, SEQ ID N°8A, SEQ ID N°9A, SEQ ID N°10A, SEQ ID N°27A ou SEQ ID N°29 contenus respectivement dans les vecteurs pDP428 (CNCM, N°I-1818), p6D7 (CNCM, N°I-1814), p5F6 (CNCM, N°I-1816), p2A29 (CNCM, N°I-1817), p5B5 (CNCM, N°I-1819), p1C7 (CNCM, N°I-1820), p2D7 (CNCM, N°I-1821), p1B7 (CNCM, N°I-1843), p5A3 (CNCM, N°I-1815) et pM1C25 (CNCM, n°I-2062).

30

L'invention concerne également un acide nucléique comprenant la totalité de la phase de lecture ouverte d'une des séquences nucléotidiques selon l'invention, notamment une des séquences SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F selon l'invention. Ledit acide nucléique peut être

35

isolé par exemple de la façon suivante :

- a) préparation d'une banque de cosmides à partir de l'ADN de *M. tuberculosis*, par exemple selon la technique décrite par Jacobs et al., 1991;
- b) hybridation de tout ou partie d'un acide nucléique sonde de séquence choisie par exemple parmi SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F avec les cosmides de la banque préalablement préparée à l'étape a) ;
- c) sélection des cosmides hybridant avec l'acide nucléique sonde de l'étape b) ;
- d) séquençage des inserts d'ADN des clones sélectionnés à l'étape c) et identification du cadre de lecture ouvert complet ;
- e) le cas échéant, clonage des inserts séquencés à l'étape d) dans un vecteur d'expression et/ou de clonage approprié.

Les acides nucléiques comprenant la totalité du cadre de lecture ouvert des séquences SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F sont parmi les acides nucléiques préférés.

La présente invention permet de déterminer un fragment de gène codant pour un polypeptide exporté. La comparaison avec la séquence du génome publiée par Cole et al. (Cole et al., 1998, Nature, 393, 537-544) permet de déterminer le gène en entier portant la séquence identifiée selon la présente invention.

Par séquence nucléotidique comprenant la totalité du cadre ouvert de lecture d'une séquence selon l'invention, on entend la séquence nucléotidique (génomique, ADNc, semi-synthétique ou synthétique) comprenant l'une des séquences

selon l'invention et s'étendant d'une part en 5' de ces séquences jusqu'au premier codon d'initiation de la traduction (ATG ou GTG) ou même jusqu'au premier codon stop, et d'autre part en 3' de ces séquences jusqu'au codon stop suivant, et ceci dans l'une quelconque des trois phases de lecture possibles.

Les séquences nucléotidiques complémentaires des séquences ci-dessus selon l'invention font également partie de l'invention.

Par polynucléotide de séquence complémentaire d'une séquence nucléotidique selon l'invention, on entend toute séquence d'ADN ou d'ARN dont les nucléotides sont complémentaires de ceux de ladite séquence selon l'invention et dont l'orientation est inversée.

Les fragments nucléotidiques des séquences ci-dessus selon l'invention notamment utiles en tant que sondes ou amorces font également partie de l'invention.

L'invention concerne aussi les polynucléotides caractérisés en ce qu'ils comprennent un polynucléotide choisi parmi :

- a) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide selon l'invention,
- b) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide selon l'invention,
- c) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide selon l'invention,
- d) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini selon l'invention.

Les conditions de forte stringence ainsi que le pourcentage d'identité seront définis ci-après dans la présente description.

Lorsque la séquence codante issue du gène marqueur d'exportation et/ou de sécrétion est une séquence issue du gène *phoA*, l'exportation et/ou la sécrétion du produit du gène *phoA*, le cas échéant tronqué, n'est obtenue que lorsque cette séquence est insérée en phase avec la séquence ou élément de régulation de l'expression de la production de polynucléotides et sa localisation placée en amont, qui contient les éléments contrôlant l'expression, l'exportation et/ou la sécrétion issus de séquence de mycobactéries.

Les vecteurs recombinants de l'invention peuvent bien entendu comprendre des sites de clonage multiples décalés de un ou deux nucléotides par rapport à un vecteur selon l'invention, permettant ainsi d'exprimer le polypeptide correspondant au fragment d'ADN de mycobactérie inséré et susceptible d'être traduit selon l'un des trois cadres de lecture possibles.

20

Par exemple les vecteurs préférés pJVEDb et pJVEDc de l'invention se distinguent du vecteur préféré pJVEDa par un décalage respectif de un et de deux nucléotides au niveau du site de clonage multiple.

25

Ainsi, les vecteurs de l'invention sont capables d'exprimer chacun des polypeptides susceptibles d'être codés par un fragment d'ADN de mycobactérie inséré. Cesdits polypeptides, caractérisés en ce qu'ils sont donc susceptibles d'être exportés et/ou sécrétés, et/ou induits ou réprimés, ou exprimés de façon constitutive lors de l'infection, font partie de l'invention.

30

On préfère notamment les polypeptides de l'invention dont les séquences d'acides aminés sont choisies parmi les séquences d'acides aminés SEQ ID N° 1 à SEQ ID N° 24C, SEQ

35

ID N° 27A à SEQ ID N° 28 et SEQ ID N° 30 à SEQ ID N° 50F et représentées respectivement par les figures 1 à 24C (planches 1 à 150), les figures 27A à 28 (planches 152 à 155) et les figures 30 à 50F (planches 157 à 275).

5

Font également partie de l'invention, les fragments ou fragments biologiquement actifs ainsi que les polypeptides homologues desdits polypeptides. Fragment, fragment biologiquement actif et polypeptides homologue de polypeptide, étant tels que définis ci-après dans la description.

L'invention concerne également les polypeptides comprenant un polypeptide ou un de leurs fragments selon l'invention.

L'invention a aussi pour objet des mycobactéries recombinantes contenant un vecteur recombinant selon l'invention décrit précédemment. Une mycobactérie préférée est une mycobactérie du type *M. smegmatis*.

M. smegmatis permet avantageusement de tester l'efficacité de séquences de mycobactéries, pour le contrôle de l'expression, de l'exportation et/ou de la sécrétion, et/ou de l'activité de promoteurs d'une séquence donnée, par exemple d'une séquence codant pour un marqueur tel que la phosphatase alcaline et/ou la luciférase.

Une autre mycobactérie préférée est une mycobactérie du type *M. bovis*, par exemple la souche BCG utilisée actuellement pour la vaccination contre la tuberculose.

Une autre mycobactérie préférée est une souche de *M. tuberculosis*, *M. bovis* ou *M. africanum* possédant potentiellement tous les systèmes de régulation appropriés.

Les inventeurs ont ainsi caractérisé en particulier un polynucléotide constitué par une séquence de nucléotides

présente chez toutes les souches testées de mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*. Ce polynucléotide, dénommé DP428 contient un cadre ouvert de lecture (ORF) codant pour un polypeptide d'environ 12 kD.

5 Le cadre de lecture ouvert (ORF) codant pour le polypeptide DP428 s'étend du nucléotide en position nt 941 au nucléotide en position nt 1351 de la séquence SEQ ID N° 2, le polypeptide DP428 ayant la séquence en acides aminés SEQ ID N° 28 suivante :

10 MKTGTATRRRLAVLIALALPGAVALLAEPSATGASDPCAASEVARTVGSVAKSMGD
YLDSPETNQVMTAVLQQQVGPGSVASLKAHFEANPKVASDLHALSQPLTDLSTRCSLP
ISGLQAIGLMQAVQGARR.

15 Ce poids moléculaire (PM) correspond au PM théorique de la protéine mature obtenue après clivage de la séquence signale, le PM de la protéine ou polypeptide DP428 étant d'environ 10 kD après ancrage potentiel au peptidoglycane et coupure potentielle entre S et G du motif LPISG.

20 Ce polynucléotide inclut, d'une part, un cadre ouvert de lecture correspondant à un gène de structure et, d'autre part, les signaux de régulation de l'expression de la séquence codante en amont et en aval de cette dernière. Le polypeptide DP428 est composé d'un peptide signal, d'une
25 région centrale hydrophile et d'une région C-terminale hydrophobe. Cette dernière se termine par deux résidus arginines (R), signal de rétention, et est précédé par un motif LPISG qui rappelle le motif LPXTG d'ancrage au peptidoglycane (Schneewind et al., 1995).

30 Par gène de structure aux fins de la présente invention, on entend un polynucléotide codant pour une protéine, un polypeptide ou encore un fragment de ces derniers, ledit polynucléotide ne comprenant que la séquence correspondant
35 au cadre ouvert de lecture (ORF), ce qui exclut les séquences du côté 5' du cadre ouvert de lecture (ORF) qui dirigent l'initiation de la transcription.

Ainsi, l'invention concerne en particulier un polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°2.

5

Plus particulièrement, l'invention concerne un polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi :

- 10 a) un polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°2,
- b) un polynucléotide dont la séquence nucléique est la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses,
- 15 de la séquence SEQ ID N°1,
- c) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide défini en a) ou b),
- d) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide défini en a), b) ou c),
- 20 e) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide défini en a), b), c) ou d),
- f) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini en a), b), c), d) ou e).

25

On entend par séquence nucléotidique, polynucléotide ou acide nucléique, selon la présente invention, aussi bien un ADN double brin, un ADN simple brin que des produits de transcription desdits ADN.

30

Par pourcentage d'identité au sens de la présente invention, on entend un pourcentage d'identité entre les bases de deux polynucléotides, ce pourcentage étant purement statistique et les différences entre les deux polynucléotides étant réparties au hasard et sur toute leur longueur.

35

Une hybridation dans des conditions de forte stringence signifie que les conditions de température et de force ionique sont choisies de telle manière qu'elles permettent le maintien de l'hybridation entre deux
5 fragments d'ADN complémentaires.

A titre illustratif, des conditions de forte stringence de l'étape d'hybridation aux fins de définir les fragments polynucléotidiques décrits ci-dessus, sont
10 avantageusement les suivantes :

l'hybridation est réalisée à une température préférentielle de 65°C, en présence de tampon commercialisé sous le nom de rapid-hyb buffer par Amersham (RPN 1636) et 100 µg/ml d'ADN de E.coli.

15 Les étapes de lavage peuvent, par exemple, être les suivantes :

- deux lavages de 10 min, préférentiellement à 65°C, dans un tampon 2 x SSC et 0,1% SDS;
- deux lavages de 10 min, préférentiellement à 65°C, dans
20 un tampon 1 x SSC et 0,1% SDS;
- un lavage de 10 min, préférentiellement à 65°C, dans un tampon de 0,1 x SSC et 0,1% SDS.

1 x SSC correspond à 0,15 M NaCl et 0,05M citrate de Na et une solution de 1 x Denhardt correspond à 0,02%
25 Ficoll, 0,02% de polyvinylpyrrolidone et 0,02% de sérum albumine bovine.

Avantageusement, un fragment nucléotidique répondant à la définition précédente aura au moins 8 nucléotides, de
30 préférence au moins 12 nucléotides, et encore plus préférentiellement au moins 20 nucléotides consécutifs de la séquence dont il est issu. Les conditions d'hybridation de forte stringence décrites ci-avant pour un polynucléotide d'une taille d'environ 200 bases, seront
35 adaptées par l'homme du métier pour des oligonucléotides de taille plus grande ou plus petite, selon l'enseignement de Sambrook et al., 1989.

Pour les conditions de mise en oeuvre des enzymes de restriction dans le but d'obtenir des fragments nucléotidiques des polynucléotides selon l'invention, on se réfèrera avantageusement à l'ouvrage de Sambrook et al., 1989.

Avantageusement, un polynucléotide de l'invention contiendra au moins une séquence comprenant l'enchaînement de nucléotides allant du nucléotide en position nt 964 au nucléotide nt 1234 du polynucléotide de séquence SEQ ID N° 1.

La présente invention a pour objet un polynucléotide selon l'invention, caractérisé en ce que sa séquence nucléique hybride avec l'ADN de séquence de mycobactéries et préférentiellement avec l'ADN de séquence de mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*.

20

Le polynucléotide est codé par une séquence polynucléotidique telle que décrite supra.

La présente invention a également pour objet un polypeptide issu d'une mycobactérie, caractérisé en ce qu'il est présent uniquement chez les mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*.

L'invention concerne également un polypeptide caractérisé en ce qu'il comprend un polypeptide choisi parmi :

a) un polypeptide dont la séquence d'acides aminés est comprise dans une séquence d'acides aminés choisie parmi les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°24C, SEQ ID N° 27A à SEQ ID N° 28 et SEQ ID N° 30 à SEQ ID N° 50F,

b) un polypeptide homologue au polypeptide défini en a),

- c) un fragment d'au moins 5 acides aminés d'un polypeptide défini en a) ou b),
d) un fragment biologiquement actif d'un polypeptide défini en a), b), ou c).

5

La présente invention a aussi pour objet un polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés SEQ ID N°1 ou SEQ ID N°2, ou un polypeptide de séquence d'acides aminés SEQ ID N°28.

10

Par polypeptide homologue, on entendra désigner les polypeptides présentant, par rapport au polypeptide naturel selon l'invention tel que le polypeptide DP428, certaines modifications comme en particulier une délétion, addition
15 ou substitution d'au moins un acide aminé, une troncation, un allongement, une fusion chimérique, et/ou une mutation. Parmi les polypeptides homologues, on préfère ceux dont la séquence d'acides aminés présente au moins 30%, de préférence 50%, d'homologie avec les séquences d'acides
20 aminés des polypeptides selon l'invention. Dans le cas d'une substitution, un ou plusieurs acides aminés consécutifs ou non consécutifs, sont remplacés par des acides aminés « équivalents ». L'expression acide aminé « équivalent » vise ici à désigner tout acide aminé
25 susceptible d'être substitué à l'un des acides aminés de la structure de base sans cependant modifier essentiellement les propriétés immunogènes des peptides correspondants. En d'autres termes, les acides aminés équivalents seront ceux qui permettent l'obtention d'un polypeptide de séquence
30 modifiée qui permet l'induction *in vivo* d'anticorps ou de cellules capables de reconnaître le polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés de polypeptide selon l'invention, telle que les séquences d'acides aminés SEQ ID N°1 à SEQ ID N° 2, ou
35 un polypeptide de séquence d'acides aminés SEQ ID N°28

(polypeptide DP428) ou l'un de ses fragments ci-dessus définis.

Ces aminoacyles équivalents peuvent être déterminés soit en s'appuyant sur leur homologie de structure avec les aminoacyles auxquels ils se substituent, soit sur les résultats des essais d'immunogénicité croisée auxquels les différents peptides sont susceptibles de donner lieu.

A titre d'exemple, on mentionnera les possibilités de substitutions susceptibles d'être effectuées sans qu'il en résulte une modification approfondie de l'immunogénicité des peptides modifiés correspondants, les remplacements, par exemple, de la leucine par la valine ou l'isoleucine, de l'acide aspartique par l'acide glutamique, de la glutamine par l'asparagine, de l'arginine par la lysine etc., les substitutions inverses étant naturellement envisageables dans les mêmes conditions.

Par fragment biologiquement actif, on entendra désigner en particulier un fragment de séquence d'acides aminés de polypeptide présentant au moins une des caractéristiques des polypeptides selon l'invention, notamment en ce qu'il est :

- capable d'être exporté et/ou sécrété par une mycobactérie, et/ou d'être induit ou réprimé lors de l'infection par la mycobactérie ; et/ou
- capable d'induire, de réprimer ou de moduler, directement ou indirectement, un facteur de virulence de mycobactérie ; et/ou
- capable d'induire une réaction d'immunogénicité dirigée contre les mycobactéries ; et/ou
- capable d'être reconnu par un anticorps spécifique de mycobactérie .

Par fragment de polypeptide, on entend désigner un polypeptide comportant au minimum 5 acides aminés, de préférence 10 acides aminés et 15 acides aminés.

Un polypeptide de l'invention, ou un de ses fragments, tels que définis précédemment, est susceptible d'être reconnu spécifiquement par les anticorps présents
5 dans le sérum de patients infectés par des mycobactéries et préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis* ou par des cellules de l'hôte infecté.

Font ainsi partie de l'invention les fragments du
10 polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés de polypeptide selon l'invention, telle que les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°2, ou un polypeptide de séquence d'acides aminés SEQ ID N°28, qui peuvent être obtenus par clivage
15 dudit polypeptide par une enzyme protéolytique, telle que la trypsine ou la chymotrypsine ou la collagénase, ou par un réactif chimique, tel que le bromure de cyanogène (CNBr) ou encore en plaçant un polypeptide selon l'invention tel que le polypeptide DP428 dans un environnement très acide,
20 par exemple à pH 2,5. Des fragments peptidiques préférés selon l'invention, pour une utilisation en diagnostic ou en vaccination, sont les fragments contenus dans des régions de polypeptide selon l'invention tel que le polypeptide DP428 susceptibles d'être naturellement exposées au solvant
25 et de présenter ainsi des propriétés d'immunogénicité importante. De tels fragments peptidiques peuvent être préparés indifféremment par synthèse chimique, à partir d'hôtes transformés par un vecteur d'expression selon l'invention contenant un acide nucléique permettant
30 l'expression desdits fragments, placé sous le contrôle des éléments de régulation et/ou d'expression appropriés ou encore par clivage chimique ou enzymatique.

Une analyse de l'hydrophilicité du polypeptide DP428 a
35 été réalisée à l'aide du logiciel DNA Strider™ (commercialisé par le CEA Saclay), sur la base d'un calcul du caractère hydrophile de la région codante pour le DP428

de la SEQ ID N°28. Les résultats de cette analyse sont présentés à la figure 54, où sont détaillés, pour chacun des acides aminés (AA) de position définie dans la SEQ ID N°28, l'indice d'hydrophilicité. Plus l'indice d'hydrophilicité est élevé, plus l'acide aminé considéré est susceptible d'être exposé au solvant dans la molécule native, et est en conséquence susceptible de présenter un degré d'antigénicité élevé. Ainsi, un enchaînement d'au moins sept acides aminés possédant un indice élevé d'hydrophilicité ($>0,3$) peut constituer la base de la structure d'un peptide candidat immunogène selon la présente invention.

Les réponses immunitaires cellulaires de l'hôte à un polypeptide selon l'invention, peuvent être mises en évidence selon les techniques décrites par Colignon et al., 1996.

D'après les données de la carte d'hydrophilicité présentée à la Figure 54, les inventeurs ont pu définir des régions du polypeptide DP428 préférentiellement exposées au solvant, plus particulièrement la région localisée entre les acides aminés 55 et 72 de la séquence SEQ ID N° 28 et la région localisée entre les acides aminés 99 et 107 de la SEQ ID N° 28.

Les régions peptidiques du polypeptide DP428 définies ci-dessus peuvent être avantageusement mises en oeuvre pour la réalisation des compositions immunogènes ou des compositions vaccinales selon l'invention.

Les polynucléotides caractérisés en ce qu'ils codent pour un polypeptide selon l'invention, font également partie de l'invention.

L'invention concerne également les séquences d'acide nucléique utilisables comme sonde ou amorce, caractérisées

en ce que lesdites séquences sont choisies parmi les séquences d'acide nucléique de polynucléotides selon l'invention.

5 L'invention concerne en outre l'utilisation d'une séquence d'acide nucléique de polynucléotides selon l'invention comme sonde ou amorce, pour la détection et/ou l'amplification de séquence d'acide nucléique. Parmi ces séquences d'acide nucléique selon l'invention utilisables
10 comme sonde ou amorce, on préfère les séquences d'acide nucléique de l'invention, caractérisée en ce que lesdites séquences sont des séquences, ou leur séquence complémentaire, comprises entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités
15 incluses, de la séquence SEQ ID N°1.

Parmi les polynucléotides selon l'invention, utilisables comme amorces nucléotidiques, on préfère particulièrement les polynucléotides de séquence SEQ ID
20 N°25 et SEQ ID N°26.

Les polynucléotides selon l'invention peuvent ainsi être utilisés pour sélectionner des amorces nucléotidiques, notamment pour la technique PCR (Erlich, 1989 ; Innis et
25 al., 1990, et, Rolfs et al., 1991).

Cette technique nécessite le choix de paires d'oligonucléotides encadrant le fragment qui doit être amplifié. On peut, par exemple, se référer à la technique
30 décrite dans le brevet américain U.S. N° 4 683 202. Ces amorces oligodésoxyribonucléotidiques ou oligoribonucléotidiques ont avantageusement une longueur d'au moins 8 nucléotides, de préférence d'au moins 12 nucléotides, et encore plus préférentiellement au moins 20 nucléotides. On
35 préférera en particulier des amorces d'une longueur comprise entre 8 et 30 et de préférence 12 et 22 nucléotides. L'une des deux amorces est complémentaires du

brin (+) [amorce aller] de la matrice et l'autre amorce est complémentaire du brin (-) [amorce retour]. Il est important que les amorces ne possèdent pas de structure secondaire ou de séquence complémentaire l'une de l'autre.

5 D'autre part, la longueur et la séquence de chaque amorce doivent être choisies de manière à ce que les amorces ne s'hybrident pas avec d'autres acides nucléiques provenant de cellules procaryotes ou eucaryotes, en particulier avec les acides nucléiques provenant d'autres mycobactéries

10 pathogènes, ni avec l'ADN ou l'ARN humain pouvant éventuellement contaminer l'échantillon biologique.

Les résultats présentés à la figure 51, montrent que la séquence codant pour le polypeptide DP428 (SEQ ID N° 28)

15 n'est pas retrouvée dans les ADNs de *M. fortuitum*, *M. simiae*, *M. avium*, *M. chelonae*, *M. flavescens*, *M. gordonae*, *M. marinum* et *M. kansasii*

Les fragments amplifiés peuvent être identifiés après

20 une électrophorèse en gel d'agarose ou de polyacrylamide, ou après une électrophorèse capillaire, ou encore après une technique chromatographique (filtration sur gel, chromatographie hydrophobe ou chromatographie échangeuse d'ions). La spécificité de l'amplification peut être

25 contrôlée par hybridation moléculaire en utilisant comme sondes les séquences nucléotidiques de polynucléotides de l'invention, des plasmides contenant ces séquences ou leurs produits d'amplification.

30 Les fragments nucléotidiques amplifiés peuvent être utilisés comme réactifs dans des réactions d'hybridation afin de mettre en évidence la présence, dans un échantillon biologique, d'un acide nucléique cible de séquence complémentaire à celle desdits fragments nucléotidiques

35 amplifiés.

Parmi les polynucléotides selon l'invention, utilisables comme sondes nucléotidiques, on préfère tout particulièrement le fragment polynucléotidique comprenant la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence de SEQ ID N°1.

Ces sondes et amplicons peuvent être marqués ou non par des éléments radioactifs ou par des molécules non radioactives, telles que des enzymes ou des éléments fluorescents..

L'invention vise également les fragments nucléotidiques susceptibles d'être obtenus par amplification à l'aide d'amorces selon l'invention.

D'autres techniques d'amplification de l'acide nucléique cible peuvent être avantageusement employées comme alternatives à la PCR.

La technique SDA (Strand Displacement Amplification) ou technique d'amplification à déplacement de brin (Walker et al., 1992) est une technique d'amplification isotherme dont le principe est fondé sur la capacité d'une enzyme de restriction de couper l'un des deux brins de son site de reconnaissance qui se trouve sous une forme hemiphosphorothioate et sur la propriété d'une ADN polymérase d'initier la synthèse d'un nouveau brin d'ADN à partir de l'extrémité 3'OH créée par l'enzyme de restriction et de déplacer le brin préalablement synthétisé qui se trouve en aval.

Les polynucléotides de l'invention, en particulier les amorces selon l'invention, peuvent également être mis en oeuvre dans d'autres procédés d'amplification d'un acide nucléique cible, tels que :

- la technique TAS (Transcription-based Amplification System), décrite par Kwoh et al. en 1989;

- la technique 3SR (Self-Sustained Sequence Replication), décrite par Guatelli et al. en 1990;
- la technique NASBA (Nucleic Acid Sequence Based Amplification), décrite par Kievitis et al. en 1991;
- 5 - la technique TMA (Transcription Mediated Amplification).

Les polynucléotides de l'invention peuvent aussi être employés dans des techniques d'amplification ou de modification de l'acide nucléique servant de sonde, telles que:

- 10 - la technique LCR (Ligase Chain Reaction), décrite par Landegren et al. en 1988 et perfectionnée par Barany et al. en 1991, qui emploie une ligase thermostable;
- la technique de RCR (Repair Chain Reaction), décrite par Segev en 1992;
- 15 - la technique CPR (Cycling Probe Reaction), décrite par Duck et al. en 1990;
- la technique d'amplification à la Q-beta-réplique, décrite par Miele et al. en 1983 et perfectionnée notamment par Chu et al. en 1986, Lizardi et al. en 1988, puis par
- 20 Burg et al. ainsi que par Stone et al. en 1996.

Dans le cas où le polynucléotide cible à détecter est un ARN, par exemple un ARNm, on utilisera avantageusement, préalablement à la mise en oeuvre d'une réaction

25 d'amplification à l'aide des amorces selon l'invention ou à la mise en oeuvre d'un procédé de détection à l'aide des sondes de l'invention, une enzyme de type transcriptase inverse afin d'obtenir un ADNc à partir de l'ARN contenu dans l'échantillon biologique. L'ADNc obtenu servira alors

30 de cible pour les amorces ou les sondes mises en oeuvre dans le procédé d'amplification ou de détection selon l'invention.

La sonde de détection sera choisie de telle manière à

35 ce qu'elle hybride avec l'amplicon généré. Une telle sonde de détection aura avantageusement pour séquence une séquence d'au moins 12 nucléotides, en particulier d'au

moins 15 nucléotides, et de préférence au moins de 200 nucléotides.

Les sondes nucléotidiques selon l'invention sont capables de détecter des mycobactéries et
5 préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis*, plus précisément du fait que ces mycobactéries possèdent dans leur génome au moins une copie de polynucléotides selon l'invention. Ces sondes selon l'invention, sont capables, par exemple, de
10 s'hybrider avec la séquence nucléotidique d'un polypeptide selon l'invention, plus particulièrement tout oligonucléotide hybridant avec la séquence SEQ ID N°1 codant pour le polypeptide DP428 de *M. tuberculosis*, et ne présentant pas de réaction d'hybridation croisée ou
15 d'amplification (PCR) avec par exemple des séquences présentes chez des mycobactéries n'appartenant pas au complexe de *Mycobacterium tuberculosis*. Les sondes nucléotidiques selon l'invention hybrident spécifiquement avec une molécule d'ADN ou d'ARN de polynucléotide selon
20 l'invention, dans des conditions d'hybridation de forte stringence telles que données sous forme d'exemple précédemment.

Les séquences non marquées peuvent être utilisées
25 directement comme sondes, cependant les séquences sont généralement marquées par un élément radioactif (^{32}P , ^{35}S , ^3H , ^{125}I) ou par une molécule non-radioactive (biotine, acétylaminofluorène, digoxigénine, 5-bromo-désoxyuridine, fluorescéine) pour obtenir des sondes utilisables pour de
30 nombreuses applications.

Des exemples de marquages non radioactifs de sondes sont décrits, par exemple, dans le brevet français N° 78.10975 ou par Urdea et al. ou par Sanchez-Pescador et al.
35 en 1988.

Dans ce dernier cas, on pourra aussi utiliser l'une des méthodes de marquage décrites dans les brevets FR 2 422 956 et FR 2 518 755. La technique d'hybridation peut être réalisée de manières diverses (Matthews et al., 1988). La
5 méthode la plus générale consiste à immobiliser l'acide nucléique extrait des cellules de mycobactéries sur un support (tel que nitrocellulose, nylon, polystyrène) et à incuber, dans des conditions bien définies, l'acide nucléique cible immobilisé avec la sonde. Après
10 l'hybridation, l'excès de sonde est éliminé et les molécules hybrides formées sont détectées par la méthode appropriée (mesure de la radioactivité, de la fluorescence ou de l'activité enzymatique liée à la sonde).

15 Avantageusement, les sondes nucléotidiques marquées selon l'invention peuvent avoir une structure telle qu'elles rendent possible une amplification du signal radioactif ou non-radioactif. Un système d'amplification répondant à la définition ci-dessus comprendra des sondes
20 de détection sous la forme d'un ADN ramifié, branché («branched DNA») telles que celles décrites par Urdea et al. en 1991. Selon cette technique, on utilisera avantageusement plusieurs types de sondes notamment une sonde de capture, afin d'immobiliser l'ADN ou l'ARN cible sur un
25 support, et une sonde de détection. La sonde de détection lie un ADN «branché» présentant une structure ramifiée. L'ADN branché, à son tour, est capable de fixer des sondes oligonucléotidiques qui sont elles-mêmes couplées à des molécules de phosphatase alcaline. Puis l'activité de cette
30 enzyme est mise en évidence grâce à un substrat chimio-luminescent, par exemple un dérivé du dioxétane-phosphate.

Selon un autre mode avantageux de mise en oeuvre des sondes nucléiques selon l'invention, ces dernières peuvent
35 être immobilisées sur un support, de manière covalente ou non covalente, et utilisées comme sondes de capture. Dans ce cas, une sonde, dite «sonde de capture», est immobilisée

sur un support et sert à capturer par hybridation spécifique l'acide nucléique cible obtenu à partir de l'échantillon biologique à tester. Si nécessaire, le support solide est séparé de l'échantillon et le duplex formé entre
5 la sonde de capture et l'acide nucléique cible est ensuite détecté grâce à une seconde sonde, dite «sonde de détection», marquée par un élément facilement détectable.

Les fragments oligonucléotidiques peuvent être obtenus
10 à partir des séquences selon l'invention, par coupure avec des enzymes de restriction, ou par synthèse chimique selon les méthodes classiques, par exemple selon la méthode décrite dans le brevet européen N° EP-0305929 (Millipore Corporation) ou encore par d'autres procédés.

15 Un mode de préparation approprié des acides nucléiques de l'invention comportant au maximum 200 nucléotides (ou 200 pb s'il s'agit d'acides nucléiques bicaténaires) comprend les étapes suivantes :

- la synthèse d'ADN en utilisant la méthode automatisée des
20 bêta-cyanethylphosphoramidite décrite en 1986,
- le clonage des acides nucléiques ainsi obtenus dans un vecteur approprié et la récupération des acides nucléiques par hybridation avec une sonde appropriée.

Un mode de préparation, par voie chimique, d'acides
25 nucléiques selon l'invention de longueur supérieure à 200 nucléotides (ou 200 pb lorsqu'il s'agit d'acides nucléiques bicaténaires) comprend les étapes suivantes :

- l'assemblage d'oligonucléotides synthétisés chimiquement, pourvus à leur extrémité de sites de restrictions
30 différents, dont les séquences sont compatibles avec l'enchaînement en acides aminés du polypeptide naturel selon le principe décrit en 1983,
- le clonage des acides nucléiques ainsi obtenus dans un vecteur approprié et la récupération des acides nucléiques
35 recherchés par hybridation avec une sonde appropriée.

Les sondes nucléotidiques utilisées pour la récupération des acides nucléiques recherchés dans les procédés sus-mentionnés, sont constituées généralement de 8 à 200 nucléotides de la séquence de polypeptide selon l'invention et sont susceptibles de s'hybrider avec l'acide nucléique recherché dans les conditions d'hybridation définies précédemment. La synthèse de ces sondes peut être effectuée selon la méthode automatisée des bêta cyanethylphosphoramidites décrite en 1986.

10

Les sondes oligonucléotidiques selon l'invention peuvent être mises en oeuvre au sein d'un dispositif de détection comprenant une banque matricielle d'oligonucléotides. Un exemple de réalisation d'une telle banque matricielle peut consister en une matrice d'oligonucléotides sondes fixés sur un support, la séquence de chaque sonde d'une longueur donnée étant située en décalage d'une ou plusieurs bases par rapport à la sonde précédente, chacune des sondes de l'arrangement matriciel étant ainsi complémentaire d'une séquence distincte de l'ADN ou l'ARN cible à détecter et chaque sonde de séquence connue étant fixée en une position prédéterminée du support. La séquence cible à détecter peut être avantageusement marquée radioactivement ou non radioactivement. Lorsque la séquence cible marquée est mise en contact avec le dispositif matriciel, celle-ci forme des hybrides avec les sondes de séquences complémentaires. Un traitement à la nucléase, suivi d'un lavage, permet d'éliminer les hybrides sondes-séquence cible qui ne sont pas parfaitement complémentaires. Du fait de la connaissance précise de la séquence d'une sonde à une position déterminée de la matrice, il est alors possible de déduire la séquence nucléotidique de la séquence d'ADN ou d'ARN cible. Cette technique est particulièrement efficace lorsque sont utilisées des matrices de sondes oligonucléotidiques de grande taille.

30

35

Une alternative à l'utilisation d'une séquence cible marquée peut consister en l'utilisation d'un support permettant une détection « bioélectronique » de l'hybridation de la séquence cible sur les sondes du support matrice, lorsque que ledit support est constitué ou comprend un matériau capable d'agir, par exemple, en tant que donneur d'électrons aux positions de la matrice auxquelles un hybride a été formé. Un tel matériau donneur d'électron est par exemple de l'or. La détection de la séquence nucléotidique de l'ADN ou ARN cible est alors déterminée par un dispositif électronique.

Un exemple de réalisation d'un biocapteur, tel que défini ci-dessus, est décrit dans la demande de brevet européen N° EP-0721 016 au nom de Affymax technologies N.V. ou encore dans le brevet américain N° US 5.202.231 au nom de Drmanac.

L'invention a aussi pour objet les polynucléotides hybrides résultant :

- soit de la formation d'une molécule hybride entre un ARN ou un ADN (ADN génomique ou ADNc) provenant d'un échantillon biologique avec une sonde ou une amorce selon l'invention.
- soit de la formation d'une molécule hybride entre un ARN ou un ADN (ADN génomique ou ADNc) provenant d'un échantillon biologique avec un fragment nucléotidique amplifié à l'aide d'un couple d'amorces selon l'invention.

Par ADNc au sens de la présente invention, on entend une molécule d'ADN obtenue en faisant agir une enzyme de type transcriptase inverse sur une molécule d'ARN, en particulier une molécule d'ARN messenger (ARNm), selon les techniques décrites dans Sambrook et al. en 1989.

La présente invention a également pour objet une famille de plasmides recombinants, caractérisés en ce qu'ils contiennent au moins une séquence nucléotidique de polynucléotide selon l'invention. Selon un mode de réalisation avantageux dudit plasmide, il comprend la

séquence nucléotidique SEQ ID N°1 ou un fragment de celle-ci.

Un autre objet de la présente invention est un vecteur pour le clonage, l'expression et/ou l'insertion d'une
5 séquence, caractérisé en ce qu'il comprend une séquence nucléotidique de polynucléotide selon l'invention en un site non essentiel pour sa répllication, le cas échéant sous le contrôle d'éléments de régulation susceptibles d'intervenir dans l'expression du polypeptide DP428, chez
10 un hôte déterminé.

Des vecteurs particuliers sont par exemple des plasmides, des phages, des cosmides, des phagemides, des YAC.

Ces vecteurs sont utiles pour transformer des cellules
15 hôtes afin de cloner ou d'exprimer les séquences nucléotidiques de l'invention.

L'invention comprend également les cellules hôtes transformées par un vecteur selon l'invention.
20

De préférence, les cellules hôtes sont transformées dans des conditions permettant l'expression d'un polypeptide recombinant selon l'invention.

25 Une cellules hôte préférée selon l'invention est la souche *E. coli* transformée par le plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818 ou transformée par le plasmide pM1C25 déposé le 4 août 1998 à la CNCM sous le n° I-2062 ou une mycobactérie appartenant à une souche
30 de *M. tuberculosis*, *M. bovis* ou *M. africanum* possédant potentiellement tous les systèmes de régulation appropriés.

Il est aujourd'hui facile de produire des protéines ou polypeptides en quantité relativement importante par génie génétique en utilisant comme vecteurs d'expression des
35 plasmides, des phages, des phagemides. Tout ou partie du gène DP428, ou tout polynucléotide selon l'invention, peut être inséré dans un vecteur d'expression approprié pour

produire *in vitro* un polypeptide selon l'invention, notamment le polypeptide DP428. Ledit polypeptide pourra être fixé sur une microplaque pour développer un test sérologique destiné à rechercher, dans un but de diagnostic, les anticorps spécifiques chez les patients atteints de tuberculose.

Ainsi, la présente invention concerne un procédé de préparation d'un polypeptide, caractérisé en ce qu'il met en oeuvre un vecteur selon l'invention. Plus particulièrement l'invention concerne un procédé de préparation d'un polypeptide de l'invention comprenant les étapes suivantes :

- le cas échéant, l'amplification préalable suivant la technique PCR de la quantité de séquences de nucléotides codant pour ledit polypeptide à l'aide de deux amorces d'ADN choisies de manière à ce que l'une de ces amorces soit identique aux 10 à 25 premiers nucléotides de la séquence nucléotidique codant pour ledit polypeptide, tandis que l'autre amorce est complémentaire des 10 à 25 derniers nucléotides (ou s'hybride avec ces 10 à 25 derniers nucléotides) de ladite séquence nucléotidique, ou inversement de manière à ce que l'une de ces amorces soit identique aux 10 à 25 derniers nucléotides de ladite séquence, tandis que l'autre amorce est complémentaire des 10 à 25 premiers nucléotides (ou s'hybride avec les 10 à 25 premiers nucléotides) de ladite séquence nucléotidique, suivie de l'introduction desdites séquences ainsi amplifiées dans un vecteur approprié,
- la mise en culture, dans un milieu de culture approprié, d'un hôte cellulaire préalablement transformé par un vecteur approprié contenant un acide nucléique selon l'invention comprenant la séquence nucléotidique codant pour ledit polypeptide, et
- la séparation, à partir du susdit milieu de culture, dudit polypeptide produit par ledit hôte cellulaire transformé.

L'invention a aussi pour objet un polypeptide susceptible d'être obtenu par un procédé de l'invention tel que décrit précédemment.

5 Les peptides selon l'invention peuvent également être préparés par les techniques classiques, dans le domaine de la synthèse des peptides. Cette synthèse peut être réalisée en solution homogène ou en phase solide.

Par exemple, on aura recours à la technique de
10 synthèse en solution homogène décrite par Houbenweyl en 1974.

Cette méthode de synthèse consiste à condenser successivement deux-à-deux les aminoacyles successifs dans
15 l'ordre requis, ou à condenser des aminoacyles et des fragments préalablement formés et contenant déjà plusieurs aminoacyles dans l'ordre approprié, ou encore plusieurs fragments préalablement ainsi préparés, étant entendu que l'on aura eu soin de protéger au préalable toutes les
20 fonctions réactives portées par ces aminoacyles ou fragments, à l'exception des fonctions amines de l'un et carboxyles de l'autre ou vice-versa, qui doivent normalement intervenir dans la formation des liaisons peptidiques, notamment après activation de la fonction
25 carboxyle, selon les méthodes bien connues dans la synthèse des peptides. En variante, on pourra avoir recours à des réactions de couplage mettant en jeu des réactifs de couplage classique, du type carbodiimide, tels que par exemple la 1-éthyl-3-(3-diméthyl-aminopropyl)-carbodiimide.

30 Lorsque l'aminoacyle mis en oeuvre possède une fonction acide supplémentaire (notamment dans le cas de l'acide glutamique), ces fonctions seront protégées, par exemple par des groupes t-butylester.

Dans le cas de la synthèse progressive, acide aminé
35 par acide aminé, la synthèse débute de préférence par la condensation de l'amino-acide C-terminal avec l'aminoacide qui correspond à l'aminoacyle voisin dans la séquence

désirée et ainsi de suite, de proche en proche, jusqu'à l'acide aminé N-terminal.

Selon une autre technique préférée de l'invention, on a recours à celle décrite par Merrifield.

5 Pour fabriquer une chaîne peptidique selon le procédé de Merrifield, on a recours à une résine polymère très poreuse, sur laquelle on fixe le premier acide aminé C-terminal de la chaîne. Cet acide aminé est fixé sur la résine par l'intermédiaire de son groupe carboxylique et sa
10 fonction amine est protégée, par exemple par le groupe t-butyloxycarbonyle.

Lorsque le premier acide aminé C-terminal est ainsi fixé sur la résine, on enlève le groupe protecteur de la
15 fonction amine en lavant la résine avec un acide.

Dans le cas où le groupe protecteur de la fonction amine est le groupe t-butyloxycarbonyle, il peut être éliminé par traitement de la résine à l'aide d'acide
20 trifluoroacétique.

On couple ensuite le deuxième acide aminé qui fournit le second aminoacyle de la séquence recherchée, à partir du résidu aminoacyle C-terminal sur la fonction amine
25 déprotégée du premier acide aminé C-terminal fixé sur la chaîne. De préférence, la fonction carboxyle de ce deuxième acide aminé est activée, par exemple par la dicyclohexylcarbodiimide, et la fonction amine est protégée, par exemple par le t-butyloxycarbonyle..

30 On obtient ainsi la première partie de la chaîne peptidique recherchée, qui comporte deux acides aminés, et dont la fonction amine terminale est protégée. Comme précédemment, on déprotège la fonction amine et on peut ensuite procéder à la fixation du troisième aminoacyle,
35 dans des conditions analogues à celles de l'addition du deuxième acide aminé C-terminal.

On fixe ainsi, les uns après les autres, les acides aminés qui vont constituer la chaîne peptidique sur le groupe amine chaque fois déprotégé au préalable de la portion de la chaîne peptidique déjà formée, et qui est
5 rattachée à la résine.

Lorsque la totalité de la chaîne peptidique désirée est formée, on élimine les groupes protecteurs des différents acides aminés constituant la chaîne peptidique et on détache le peptide de la résine, par exemple à l'aide
10 d'acide fluorhydrique.

De manière préférentielle, lesdits polypeptides susceptibles d'être obtenus par un procédé de l'invention tel que décrit précédemment comprendront une région exposée
15 au solvant et auront une longueur d'au moins 20 acides aminés.

Selon un autre mode de réalisation de l'invention, lesdits polypeptides sont spécifiques de mycobactéries du complexe *Mycobacterium tuberculosis* et ne sont donc pas
20 reconnus par des anticorps spécifiques d'autres protéines de mycobactéries.

L'invention est en outre relative à des polypeptides
25 hybrides présentant au moins un polypeptide selon l'invention et une séquence d'un polypeptide susceptible d'induire une réponse immunitaire chez l'homme ou l'animal.

Avantageusement, le déterminant antigénique est tel qu'il est susceptible d'induire une réponse humorale et/ou
30 cellulaire.

Un tel déterminant pourra comprendre un polypeptide selon l'invention sous forme glycosylée utilisé en vue d'obtenir des compositions immunogènes susceptibles
35 d'induire la synthèse d'anticorps dirigés contre des épitopes multiples. Lesdits polypeptides glycosylés font également partie de l'invention.

Ces molécules hybrides peuvent être constituées en partie d'une molécule porteuse de polypeptide selon l'invention associée à une partie, en particulier un épitope de la toxine diphtérique, la toxine tétanique, un
5 antigène de surface du virus de l'hépatite B (brevet FR 79 21811), l'antigène VP1 du virus de la poliomyélite ou toute autre toxine ou antigène viral ou bactérien.

Avantageusement, ledit déterminant antigénique
10 correspond à un déterminant antigénique de protéines immunogènes de 45/47 kD de *M. tuberculosis* (demande internationale PCT/FR 96/0166), ou encore sélectionnées par exemple parmi ESAT6 (Harboe et al., 1996, Andersen et al., 1995, et Sorensen et al., 1995) et DES (PCT/FR 97/00923,
15 Gicquel et al.).

Un antigène viral, tel que défini ci-dessus, sera préférentiellement une protéine de surface ou d'enveloppe d'un virus de l'hépatite, par exemple la protéine de
20 surface de l'hépatite B sous l'une de ses formes S, S-prés1, S-prés2 ou S-prés2-prés1 ou encore une protéine d'un virus de l'hépatite A, ou d'une hépatite non-A, non-B, tel qu'un virus de l'hépatite C, E ou delta.

25 Plus particulièrement, un antigène viral tel que défini ci-dessus sera tout ou partie de l'une des glycoprotéines codées par le génome du virus HIV-1 (brevets GB 8324800, EP 84401834 ou EP 85905513) ou du virus HIV-2 (EP 87400151), et en particulier tout ou partie d'une
30 protéine sélectionnée parmi gag, pol, nef ou env de HIV-1 ou de HIV-2.

Les procédés de synthèse des molécules hybrides englobent les méthodes utilisées en génie génétique pour
35 construire des polynucléotides hybrides codant pour les séquences polypeptidiques recherchées. On pourra, par exemple, se référer avantageusement à la technique

d'obtention de gènes codant pour des protéines de fusion décrite par Minton en 1984.

Lesdits polynucléotides hybrides codant pour un polypeptide hybride ainsi que les polypeptides hybrides selon l'invention caractérisés en ce qu'il s'agit de protéines recombinantes obtenues par l'expression desdits polynucléotides hybrides, font également partie de l'invention.

10

Les polypeptides selon l'invention peuvent avantageusement être mis en oeuvre dans un procédé pour la détection *in vitro* d'anticorps dirigés contre lesdits polypeptides, notamment le polypeptide DP428, et ainsi d'anticorps dirigés contre une bactérie du complexe *Mycobacterium tuberculosis*, dans un échantillon biologique (tissu ou fluide biologique) susceptible de les contenir, ce procédé comprenant la mise en contact de cet échantillon biologique avec un polypeptide selon l'invention dans des conditions permettant une réaction immunologique *in vitro* entre ledit polypeptide et les anticorps éventuellement présents dans l'échantillon biologique, et la mise en évidence *in vitro* des complexes antigène-anticorps éventuellement formés.

25

Les polypeptides selon l'invention peuvent également et avantageusement être mis en oeuvre dans un procédé pour la détection d'une infection par une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère basé sur la détection *in vitro* d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide comme par exemple la prolifération cellulaire, la synthèse de protéines telles que l'interféron gamma. Ce procédé pour la détection d'une infection par une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère, est caractérisé en ce qu'il comprend les étapes suivantes :

35

- a) préparation d'un échantillon biologique contenant des cellules dudit mammifère plus particulièrement des cellules du système immunitaire dudit mammifère et plus particulièrement encore des cellules T ;
- 5 b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'invention;
- c) détection d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide comme par exemple la prolifération cellulaire et/ou la
- 10 synthèse de protéines telles que l'interféron gamma.
- La prolifération cellulaire pourra être mesurée, par exemple par incorporation de ³H-Thymidine.

Font également partie de l'invention, les procédés de

15 détection d'une réaction d'hypersensibilité retardée (DTH), caractérisés en ce qu'ils mettent en oeuvre un polypeptide selon l'invention.

De préférence, l'échantillon biologique est constitué

20 par un fluide, par exemple un sérum humain ou animal, du sang, des biopsies, le liquide broncho-alvéolaire ou le liquide pleural.

Toute procédure classique peut être mise en oeuvre

25 pour réaliser une telle détection.

A titre d'exemple, une méthode préférée met en jeu des processus immunoenzymatiques selon la technique ELISA, par immunofluorescence, ou radio-immunologique (RIA) ou

30 équivalent.

Ainsi, l'invention concerne également les polypeptides selon l'invention, marqués à l'aide d'un marqueur adéquat tel que du type enzymatique, fluorescent, radioactif.

35 De telles méthodes comprennent par exemple les étapes suivantes :

- dépôt de quantités déterminées d'une composition polypeptidique selon l'invention dans les puits d'une plaque de microtitration,
- introduction dans lesdits puits de dilutions croissantes de sérum, ou d'échantillon biologique autre tel que défini précédemment, devant être analysé,
- incubation de la microplaque,
- introduction dans les puits de la plaque de microtitration d'anticorps marqués dirigés contre des immunoglobulines humaines ou animales, le marquage de ces anticorps ayant été réalisé à l'aide d'une enzyme sélectionnée parmi celles qui sont capables d'hydrolyser un substrat en modifiant l'absorption des radiations de ce dernier, au moins à une longueur d'onde déterminée, par exemple à 550 nm,
- détection, en comparaison avec un témoin de contrôle, de la quantité de substrat hydrolysé.

L'invention concerne également un nécessaire ou kit pour le diagnostic in vitro d'une infection par une mycobactérie appartenant au complexe *Mycobacterium tuberculosis*, comprenant:

- un polypeptide selon l'invention,
- le cas échéant les réactifs pour la constitution du milieu propice à la réaction immunologique ou spécifique,
- les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique éventuellement présents dans l'échantillon biologique, et la mise en évidence in vitro des complexes antigène-anticorps éventuellement formés, ces réactifs pouvant également porter un marqueur, ou être susceptibles d'être reconnus à leur tour par un réactif marqué, plus particulièrement dans le cas où le polypeptide selon l'invention n'est pas marqué,
- le cas échéant, un échantillon biologique de référence (témoin négatif) dépourvu d'anticorps reconnus par un polypeptide selon l'invention,

- le cas échéant, un échantillon biologique de référence (témoin positif) contenant une quantité prédéterminée d'anticorps reconnus par un polypeptide selon l'invention.

5 Les polypeptides selon l'invention permettent de préparer des anticorps monoclonaux ou polyclonaux caractérisés en ce qu'ils reconnaissent spécifiquement les polypeptides selon l'invention. Les anticorps monoclonaux pourront avantageusement être préparés à partir
10 d'hybridomes selon la technique décrite par Kohler et Milstein en 1975. Les anticorps polyclonaux pourront être préparés, par exemple par immunisation d'un animal, en particulier une souris, avec un polypeptide selon l'invention associé à un adjuvant de la réponse
15 immunitaire, puis purification des anticorps spécifiques contenus dans le sérum des animaux immunisés sur une colonne d'affinité sur laquelle a préalablement été fixé le polypeptide ayant servi d'antigène. Les anticorps polyclonaux selon l'invention peuvent aussi être préparés
20 par purification sur une colonne d'affinité, sur laquelle a préalablement été immobilisé un polypeptide selon l'invention, des anticorps contenus dans le sérum de patients infectés par une mycobactérie et préférentiellement une bactérie appartenant au complexe
25 *Mycobacterium tuberculosis*.

L'invention a également pour objet des anticorps mono ou polyclonaux ou leurs fragments, ou anticorps
30 chimériques, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'invention.

Les anticorps de l'invention pourront également être
35 marqués de la même manière que décrit précédemment pour les

sondes nucléiques de l'invention tel qu'un marquage de type enzymatique, fluorescent ou radioactif.

L'invention vise en outre un procédé pour la détection
5 spécifique de la présence d'un antigène d'une mycobactérie
et préférentiellement un bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en
ce qu'il comprend les étapes suivantes :

- 10 a) Mise en contact de l'échantillon biologique (tissu ou fluide biologique) prélevé chez un individu avec un anticorps mono ou polyclonal selon l'invention , dans des conditions permettant une réaction immunologique in vitro
entre lesdits anticorps et les polypeptides spécifiques des
15 mycobactéries et préférentiellement des bactéries du
complexe de *Mycobacterium tuberculosis* éventuellement
présents dans l'échantillon biologique, et
b) Mise en évidence du complexe antigène-anticorps formé.

20 Entre également dans le cadre de l'invention, un
nécessaire ou kit pour le diagnostic in vitro sur un
échantillon biologique, de la présence de souches de
mycobactéries des mycobactéries et préférentiellement des
bactéries appartenant au complexe de *Mycobacterium*
25 *tuberculosis*, de préférence *M. tuberculosis*, caractérisé en
ce qu'il comprend :

- un anticorps polyclonal ou monoclonal selon l'invention,
le cas échéant marqué;
- le cas échéant, un réactif pour la constitution du milieu
30 propice à la réalisation de la réaction immunologique;
- un réactif permettant la détection des complexes
antigène-anticorps produits par la réaction immunologique,
ce réactif pouvant également porter un marqueur, ou être
susceptible d'être reconnu à son tour par un réactif
35 marqué, plus particulièrement dans le cas où ledit
anticorps monoclonal ou polyclonal n'est pas marqué.

- le cas échéant, des réactifs pour effectuer la lyse des cellules de l'échantillon testé.

La présente invention a également pour objet un
5 procédé de détection et d'identification rapide des mycobactéries et préférentiellement des bactéries de *M. tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comporte les étapes suivantes :

- a) Isolement de l'ADN à partir de l'échantillon
10 biologique à analyser, ou obtention d'un ADNc à partir de l'ARN de l'échantillon biologique ;
- b) Amplification spécifique de l'ADN des mycobactéries et préférentiellement des bactéries appartenant au complexe *Mycobacterium tuberculosis* à l'aide d'amorces selon
15 l'invention;
- c) Analyse des produits d'amplification.

Le produits d'amplification peuvent être analysés par différentes méthodes.

20 Deux méthodes d'analyse sont données à titre d'exemple ci-dessous :

- Analyse électrophorétique en gel d'agarose des produits d'amplification. La présence d'un fragment d'ADN migrant à l'endroit attendu suggère que l'échantillon
25 analysé contenait de l'ADN de mycobactéries appartenant au complexe tuberculosis, ou

- Analyse par la technique d'hybridation moléculaire en utilisant une sonde nucléique selon l'invention. Cette sonde sera avantageusement marquée par un élément non
30 radioactif (sonde froide) ou radioactif.

Aux fins de la présente invention, on entendra par « ADN de l'échantillon biologique » ou « ADN contenu dans l'échantillon biologique », soit l'ADN présent dans
35 l'échantillon biologique considéré, soit l'ADNc obtenu après l'action d'une enzyme de type transcriptase inverse sur l'ARN présent dans ledit échantillon biologique.

Un autre procédé de la présente invention permet la détection d'une infection par une mycobactérie et
5 préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère. Ce procédé comprend les étapes suivantes :

- a) préparation d'un échantillon biologique contenant des cellules dudit mammifère plus particulièrement des cellules
10 du système immunitaire dudit mammifère et plus particulièrement encore des cellules T ;
- b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'invention;
- c) détection d'une réaction cellulaire indiquant une
15 sensibilisation préalable du mammifère audit polypeptide notamment la prolifération cellulaire et/ou la synthèse de protéines telles que l'interféron gamma;
- d) détection d'une réaction d'hypersensibilité retardée ou de sensibilisation du mammifère audit polypeptide.

20 Cette méthode de détection est une méthode intradermique, qui est décrite par exemple par M. J. Elhay et al. (1988) *Infection and Immunity*, 66(7) : 3454-3456.

25 Un autre but de la présente invention consiste en un procédé pour la détection des mycobactéries et préférentiellement des bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- 30 a) Mise en contact d'une sonde oligonucléotidique selon l'invention avec un échantillon biologique, l'ADN contenu dans l'échantillon biologique, ou l'ADNc obtenu par transcription inverse de l'ARN de l'échantillon biologique, ayant, le cas échéant, préalablement été rendu accessible à
35 l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN ou l'ADNc des mycobactéries et

préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;

b) Détection de l'hybride formé entre la sonde oligonucléotidique et l'ADN de l'échantillon biologique.

5

L'invention vise également un procédé pour la détection des mycobactéries et préférentiellement des bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

10

a) Mise en contact d'une sonde oligonucléotidique selon l'invention immobilisée sur un support, avec un échantillon biologique, l'ADN de l'échantillon biologique ayant, le cas échéant, été préalablement rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de ladite sonde à l'ADN des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;

15

b) Mise en contact de l'hybride formé entre ladite sonde oligonucléotidique immobilisée sur un support et l'ADN contenu dans l'échantillon biologique, le cas échéant après élimination de l'ADN de l'échantillon biologique n'ayant pas hybridé avec la sonde, avec une sonde oligonucléotidique marquée selon l'invention.

20

Selon un mode de réalisation avantageux du procédé de détection défini précédemment, celui-ci est caractérisé en ce que, préalablement à l'étape a), l'ADN de l'échantillon biologique est préalablement amplifié à l'aide d'un couple d'amorces selon l'invention.

25

Une autre forme de mise en oeuvre du procédé de détection selon l'invention consiste en un procédé pour la détection de la présence des mycobactéries et préférentiellement des bactéries appartenant au complexe de

30

Mycobacterium tuberculosis dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

5 a) Mise en contact de l'échantillon biologique avec un couple d'amorces selon l'invention, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation desdites amorces à l'ADN des mycobactéries et préférentiellement des bactéries du
10 complexe *Mycobacterium tuberculosis*;

b) Amplification de l'ADN d'une mycobactérie et préférentiellement d'une bactérie du complexe *Mycobacterium tuberculosis*;

15 c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique selon invention.

20 L'invention a aussi pour objet un procédé pour la détection de la présence des mycobactéries et préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique
25 par déplacement de brin, caractérisé en ce qu'il comprend les étapes suivantes :

a) Mise en contact de l'échantillon biologique avec deux couples d'amorces selon l'invention spécifiquement destinées à l'amplification de type SDA décrites ci-dessus,
30 l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;

- b) amplification de l'ADN des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique selon l'invention.

L'invention concerne aussi un nécessaire ou kit pour la mise en oeuvre du procédé décrit ci-dessus, destiné à la détection de la présence des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

- a) Une sonde oligonucléotidique selon l'invention;
- b) Les réactifs nécessaires à la mise en oeuvre d'une réaction d'hybridation;
- c) Le cas échéant, un couple d'amorces selon l'invention ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN (ADN génomique, ADN plasmidique ou ADNc) des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*.

L'invention a aussi pour objet un kit ou nécessaire pour la détection de la présence des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

- a) Une sonde oligonucléotidique, dite sonde de capture, selon l'invention;
- b) Une sonde oligonucléotidique, dite sonde de révélation, selon l'invention.
- c) Le cas échéant, un couple d'amorces selon l'invention ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN des mycobactéries et

préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*.

L'invention concerne encore un kit ou nécessaire pour
5 l'amplification de l'ADN des mycobactéries et
préférentiellement des bactéries du complexe *Mycobacterium tuberculosis* présent dans un échantillon biologique,
caractérisé en ce qu'il comprend les éléments suivants :

- 10 a) Un couple d'amorces selon l'invention;
- b) Les réactifs nécessaires pour effectuer une réaction d'amplification d'ADN;
- c) Eventuellement un composant permettant de vérifier la
séquence du fragment amplifié, plus particulièrement une
15 sonde oligonucléotidique selon l'invention.

Un autre objet de la présente invention concerne une
composition immunogène, caractérisée en ce qu'elle comprend
un polypeptide selon l'invention.

20 Une autre composition immunogène selon l'invention est
caractérisé en ce qu'elle comprend un ou plusieurs
polypeptides selon l'invention et/ou un ou plusieurs
polypeptides hybrides selon l'invention.

25 Selon un mode de réalisation avantageux, la
composition immunogène ci-dessus définie est constitutive
d'un vaccin, lorsqu'elle est présentée en association avec
un véhicule pharmaceutiquement acceptable et éventuellement
30 un ou plusieurs adjuvants de l'immunité tels que l'alun ou
un représentant de la famille des muramyl peptides ou
encore l'adjuvant incomplet de Freund.

Aujourd'hui, divers types de vaccins sont disponibles
35 pour protéger l'homme contre des maladies infectieuses :
micro-organismes vivants atténués (*M. bovis* - BCG pour la

tuberculose), micro-organismes inactivés (virus de la grippe), des extraits acellulaires (*Bordetella pertussis* pour la coqueluche), protéines recombinées (antigène de surface du virus de l'hépatite B), des polysides (pneumocoques). Des vaccins préparés à partir de peptides de synthèse ou de micro-organismes génétiquement modifiés exprimant des antigènes hétérologues sont en cours d'expérimentation. Plus récemment encore, des ADN plasmidiques recombinés portant des gènes codant pour des antigènes protecteurs ont été proposés comme stratégie vaccinale alternative. Ce type de vaccination est réalisé avec un plasmide particulier dérivant d'un plasmide de *E. coli* qui ne se réplique pas *in vivo* et qui code uniquement pour la protéine vaccinnante. Les principaux composants fonctionnels de ce plasmide sont : un promoteur fort permettant l'expression dans les cellules eucaryotes (par exemple celui du CMV), un site de clonage approprié pour insérer le gène d'intérêt, une séquence de terminaison-polyadénylation, une origine de réplication procaryote pour produire le plasmide recombiné *in vitro* et un marqueur de sélection (par exemple le gène de résistance à l'ampicilline) pour faciliter la sélection des bactéries qui contiennent le plasmide. Des animaux ont été immunisés en injectant simplement l'ADN plasmidique nu dans le muscle. Cette technique conduit à l'expression de la protéine vaccinnale *in situ* et à une réponse immunitaire en particulier de type cellulaire (CTL) et de type humoral (anticorps). Cette double induction de la réponse immunitaire est l'un des principaux avantages de la technique de vaccination avec de l'ADN nu. Huygen et al. (1996) et Tascon et al. (1996) ont réussi à obtenir une certaine protection contre *M. tuberculosis* en injectant des plasmides recombinés contenant des gènes de *M. leprae* (*hsp65*, *36kDa pra*) comme inserts. *M. leprae* est l'agent responsable de la lèpre. L'utilisation d'un insert spécifique de *M. tuberculosis* comme par exemple

tout ou partie du gène DP428, objet de la présente invention conduirait probablement à une meilleure protection contre la tuberculose. Tout ou partie du gène DP428, ou tout polynucléotide selon l'invention, peut être
5 facilement inséré dans les plasmides vecteurs V1J (Montgomery et al, 1993), pcDNA3 (Invitrogen, R & D Systems) ou pcDNA1/Neo (Invitrogen) qui possèdent les caractéristiques nécessaires pour une utilisation vaccinale.

10

L'invention vise ainsi un vaccin, caractérisée en ce qu'il comprend un ou plusieurs polypeptides selon l'invention et/ou un ou plusieurs polypeptides hybrides selon l'invention tels que précédemment définis en
15 association avec un véhicule pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité appropriés.

L'invention vise aussi une composition vaccinale destinée à l'immunisation de l'homme ou l'animal à
20 l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, caractérisée en ce qu'elle comprend un ou plusieurs polypeptides hybrides tels que précédemment définis en association avec un véhicule
25 pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité.

Avantageusement, dans le cas d'une protéine hybride entre un polypeptide selon l'invention et l'antigène de
30 surface de l'hépatite B, la composition vaccinale sera administrée, chez l'homme, à raison de 0,1 à 1 μ g de protéine hybride purifiée par kilogramme du poids du patient, de préférence 0,2 à 0,5 μ g/kg de poids du patient, pour une dose destinée à une administration donnée. Dans le
35 cas de patients atteints de troubles du système immunitaire, en particulier les patients immunodéprimés, chaque dose injectée contiendra préférentiellement la

moitié de la quantité pondérale de la protéine hybride contenue dans une dose destinée à un patient n'étant pas affecté de troubles du système immunitaire.

5 De préférence, la composition vaccinale sera administrée à plusieurs reprises, de manière étalée dans le temps, par voie intradermique ou sous-cutanée. A titre d'exemple, trois doses telles que définies ci-dessus seront respectivement administrées au patient au temps t_0 , au
10 temps $t_0 + 1$ mois et au temps $t_0 + 1$ an.

Alternativement, trois doses seront respectivement administrées au patient au temps t_0 , au temps $t_0 + 1$ mois et au temps $t_0 + 6$ mois.

15 Chez la souris, chez laquelle une dose pondérale de la composition vaccinale comparable à la dose utilisée chez l'homme est administrée, la réaction anticorps est testée par prélèvement du sérum suivi d'une étude de la formation d'un complexe entre les anticorps présents dans le sérum et
20 l'antigène de la composition vaccinale, selon les techniques usuelles.

L'invention concerne également une composition immunogène caractérisée en ce qu'elle comprend un
25 polynucléotide ou un vecteur d'expression selon l'invention, en association avec un véhicule permettant son administration à l'homme ou l'animal.

L'invention a encore pour objet un vaccin destiné à
30 l'immunisation à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, caractérisé en ce qu'il comprend un polynucléotide ou un vecteur d'expression selon l'invention, en association avec un véhicule pharmaceutiquement acceptable.

35

De telles compositions immunogènes ou vaccinales sont notamment décrites dans la demande internationale N° WO

90/11092 (Vical Inc.) et également dans la demande internationale N° WO 95/11307 (Institut Pasteur).

Le polynucléotide constitutif de la composition immunogène ou de la composition vaccinale selon l'invention peut être injecté à l'hôte après avoir été couplé à des composés qui favorisent la pénétration de ce polynucléotide à l'intérieur de la cellule ou son transport jusqu'au noyau cellulaire. Les conjugués résultants peuvent être encapsulés dans des microparticules polymères, comme décrit dans la demande internationale N° WO 94/27238 (medisorb Technologies International).

Selon un autre mode de réalisation de la composition immunogène et/ou vaccinale selon l'invention, le polynucléotide, de préférence un ADN, est complexé avec du DEAE-dextran (Pagano et al., 1967) ou avec des protéines nucléaires (Kaneda et al., 1989), avec des lipides (Felgner et al., 1987) ou encore encapsulés dans des liposomes (Fraley et al., 1980).

Selon encore un autre mode de réalisation avantageux de la composition immunogène et/ou vaccinale selon l'invention, le polynucléotide selon l'invention peut être introduit sous la forme d'un gel facilitant sa transfection dans les cellules. Une telle composition sous forme de gel peut être un complexe de poly-L-lysine et de lactose, comme décrit par Midoux en 1993, ou encore le Poloxamer 407™, comme décrit par Pastore en 1994. Le polynucléotide ou le vecteur selon l'invention peuvent aussi être en suspension dans une solution tampon ou être associés à des liposomes.

Avantageusement, un tel vaccin sera préparé conformément à la technique décrite par Tacson et al. ou Huygen et al. en 1996 ou encore conformément à la technique décrite par Davis et al. dans la demande internationale N° WO 95/11307 (Whalen et al.).

Un tel vaccin sera avantageusement préparé sous la forme d'une composition contenant un vecteur selon l'invention, placée sous le contrôle d'éléments de régulation permettant son expression chez l'homme ou l'animal.

Pour réaliser un tel vaccin, le polynucléotide selon l'invention est tout d'abord sous-cloné dans un vecteur d'expression approprié, plus particulièrement un vecteur d'expression contenant des signaux de régulation et d'expression reconnus par les enzymes des cellules eucaryotes et contenant également une origine de répllication active chez les procaryotes, par exemple chez *E. coli*, qui permet son amplification préalable. Puis le plasmide recombinant purifié obtenu est injecté à l'hôte, par exemple par voie intramusculaire.

On pourra par exemple utiliser, en tant que vecteur d'expression in vivo de l'antigène d'intérêt, le plasmide pcDNA3 ou le plasmide pcDNA1/neo, tous les deux commercialisés par Invitrogen (R&D Systems, Abingdon, Royaume-Uni). On peut aussi utiliser le plasmide V1Jns.tPA, décrit par Shiver et al. en 1995.

Un tel vaccin comprendra avantageusement, outre le vecteur recombinant, une solution saline, par exemple une solution de chlorure de sodium.

Une composition vaccinale telle que définie ci-dessus sera par exemple administrée par voie parentérale ou par voie intramusculaire.

La présente invention concerne également un vaccin caractérisé en ce qu'il contient une ou plusieurs séquences nucléotidiques selon l'invention et/ou un ou plusieurs polynucléotides tel que mentionné ci-dessus en association avec un véhicule pharmaceutiquement compatible et, le cas

échéant, un ou plusieurs adjuvants de l'immunité appropriés.

Un autre aspect porte sur une méthode de criblage de
5 molécules capables d'inhiber la croissance de mycobactéries
ou le maintien de mycobactéries dans un hôte, caractérisée
en ce que lesdites molécules bloquent la synthèse ou la
fonction des polypeptides codés par une séquence
nucléotidique selon l'invention ou par un polynucléotide
10 tel que décrit supra.

Dans ladite méthode de criblage, les molécules peuvent
être des anti-messagers ou peuvent induire la synthèse
d'anti-messagers.

15

La présente invention vise également des molécules
capables d'inhiber la croissance de mycobactéries ou le
maintien de mycobactéries dans un hôte, caractérisées en ce
que lesdites molécules sont synthétisées d'après la
20 structure des polypeptides codés par une séquence
nucléotidique selon l'invention ou par un polynucléotide
tel que décrit supra.

D'autres caractéristiques et avantages de l'invention
25 apparaissent dans les exemples et les figures suivants :

FIGURES

30 La série de Figures 1 :

La série de Figures 1 illustre la série de séquences
nucléotidiques SEQ ID N°1 correspondant à l'insert du
vecteur pDP428 (déposé à la CNCM sous le N° I-1818) et
35 la série de séquences d'acides aminés SEQ ID N°1 des

polypeptides codés par la série des séquences nucléotidiques SEQ ID N°1.

Figure 2 :

- 5 Illustre la séquence nucléotidique SEQ ID N°2 correspondant à la région incluant le gène codant pour le polypeptide DP428 (région soulignée). Sur cette figure ont été pris en compte à la fois les codons ATG et GTG d'initiation de la traduction. La figure fait apparaître que le polypeptide DP428 fait probablement partie d'un opéron comprenant au moins trois gènes. La région doublement encadrée inclut probablement les régions promotrices.
- 10 La région simplement encadrée correspond au motif LPISG rappelant le motif LPXTG décrit chez les bactéries à Gram positifs comme permettant l'ancrage aux peptidoglycannes.
- 15

20 La série de Figures 3 :

La série de Figures 3 représente la série de séquences nucléotidiques SEQ ID N°3 correspondant à l'insert du vecteur p6D7 (déposé à la CNCM sous le N° I-1814).

25

La série de Figures 4 :

La série de Figures 4 représente la série de séquences nucléotidiques SEQ ID N°4 correspondant à l'insert du vecteur p5A3 (déposé à la CNCM sous le N° I-1815).

30

La série de Figures 5 :

La série de Figures 5 représente la série de séquences nucléotidiques SEQ ID N°5 correspondant à l'insert du vecteur p5F6 (déposé à la CNCM sous le N° I-1816).

35

La série de Figures 6 :

La série de Figures 6 représente la série de séquences nucléotidiques SEQ ID N°6 correspondant à l'insert du vecteur p2A29 (déposé à la CNCM sous le N° I-1817).

La série de Figures 7 :

La série de Figures 7 représente la série de séquences nucléotidiques SEQ ID N°7 correspondant à l'insert du vecteur p5B5 (déposé à la CNCM sous le N° I-1819).

La série de Figures 8 :

La série de Figures 8 représente série de séquences nucléotidiques SEQ ID N°8 correspondant à l'insert du vecteur p1C7 (déposé à la CNCM sous le N° I-1820).

La série de Figures 9 :

La série de Figures 9 représente la série de séquences nucléotidiques SEQ ID N°9 correspondant à l'insert du vecteur p2D7 (déposé à la CNCM sous le N° I-1821).

La série de Figures 10 :

La série de Figures 10 représente la série de séquences nucléotidiques SEQ ID N°10 correspondant à l'insert du vecteur p1B7 (déposé à la CNCM sous le N° I-1843).

La série de Figures 11 :

La série de Figures 11 représente la série de séquences nucléotidiques SEQ ID N°11.

La série de Figures 12 :

La série de Figures 12 représente la série de séquences nucléotidiques SEQ ID N°12.

La série de Figures 13 :

5

La série de Figures 13 représente la série de séquences nucléotidiques SEQ ID N°13.

La série de Figures 14 :

10

La série de Figures 14 représente la série de séquences nucléotidiques SEQ ID N°14 correspondant à l'insert du vecteur p5B5 (déposé à la CNCM sous le N° I-1819).

15

La série de Figures 15 :

20 La série de Figures 15 représente la série de séquences nucléotidiques SEQ ID N°15.

La série de Figures 16 :

25 La série de Figures 16 représente la série de séquences nucléotidiques SEQ ID N°16.

La série de Figures 17 :

30 La série de Figures 17 représente la série de séquences nucléotidiques SEQ ID N°17.

La série de Figures 18 :

35 La série de Figures 18 représente la série de séquences nucléotidiques SEQ ID N°18.

La série de Figures 19 :

La série de Figures 19 représente la série de séquences nucléotidiques SEQ ID N°19.

5 La série de Figures 20 :

La série de Figures 20 représente la série de séquences nucléotidiques SEQ ID N°20 correspondant à l'insert du vecteur p2A29 (déposé à la CNCM sous le N° I-1817).

10

La série de Figures 21 :

La série de Figures 21 représente la série de séquences nucléotidiques SEQ ID N°21.

15

La série de Figures 22 :

La série de Figures 22 représente la série de séquences nucléotidiques SEQ ID N°22.

20

La série de Figures 23 :

La série de Figures 23 représente la série de séquences nucléotidiques SEQ ID N°23.

25

La série de Figures 24 :

La série de Figures 24 représente la série de séquences nucléotidiques SEQ ID N°24.

30

Figures 25 et 26 :

Les figures 25 et 26 illustrent respectivement les séquences SEQ ID N°25 et SEQ ID N°26 représentant un couple d'amorces utilisées pour amplifier spécifiquement par PCR la région correspondant aux nucléotides 964 à 1234 inclus dans la séquence SEQ ID N°1.

35

La série de Figures 27 :

- 5 La série de Figures 27 représente la série de séquences nucléotidiques SEQ ID N°27 correspondant à l'insert du vecteur p5A3.

Figure 28 :

- 10 La séquence d'acides aminés telle que définie dans la figure 28 représente la séquence d'acides aminés SEQ ID N°28 correspondant au polypeptide DP428.

15 Figure 29 :

La figure 29 représente la séquence nucléotidique SEQ ID N° 29 du gène complet codant pour la protéine M1C25.

20 Figure 30 :

La figure 30 représente la séquence d'acides aminés SEQ ID N° 30 de la protéine M1C25.

25 La série de Figures 31 :

La série de Figures 31 représente la série de séquences nucléotidiques SEQ ID N°31.

30 La série de Figures 32 :

La série de Figures 32 représente la série de séquences nucléotidiques SEQ ID N°32.

35 La série de Figures 33 :

La série de Figures 33 représente la série de séquences nucléotidiques SEQ ID N°33.

La série de Figures 34 :

La série de Figures 32 représente la série de séquences nucléotidiques SEQ ID N°34.

La série de Figures 35 :

La série de Figures 35 représente la série de séquences nucléotidiques SEQ ID N°35.

La série de Figures 36 :

La série de Figures 36 représente la série de séquences nucléotidiques SEQ ID N°36.

La série de Figures 37 :

La série de Figures 37 représente la série de séquences nucléotidiques SEQ ID N°37.

La série de Figures 38 :

La série de Figures 38 représente la série de séquences nucléotidiques SEQ ID N°38.

La série de Figures 39 :

La série de Figures 39 représente la série de séquences nucléotidiques SEQ ID N°39.

La série de Figures 40 :

La série de Figures 40 représente la série de séquences nucléotidiques SEQ ID N°40.

La série de Figures 41 :

5

La série de Figures 41 représente la série de séquences nucléotidiques SEQ ID N°41 correspondant à l'insert du vecteur p2D7 (déposé à la CNCM sous le N°I-1821).

10 La série de Figures 42 :

La série de Figures 42 représente la série de séquences nucléotidiques SEQ ID N°42.

15

La série de Figures 43 :

20 La série de Figures 43 représente la série de séquences nucléotidiques SEQ ID N°43.

La série de Figures 44 :

25 La série de Figures 44 représente la série de séquences nucléotidiques SEQ ID N°44.

La série de Figures 45 :

30 La série de Figures 45 représente la série de séquences nucléotidiques SEQ ID N°45.

La série de Figures 46 :

35 La série de Figures 46 représente la série de séquences nucléotidiques SEQ ID N°46.

La série de Figures 47 :

La série de Figures 47 représente la série de séquences nucléotidiques SEQ ID N°47.

5 La série de Figures 48 :

La série de Figures 48 représente la série de séquences nucléotidiques SEQ ID N°48.

10 La série de Figures 49 :

La série de Figures 49 représente la série de séquences nucléotidiques SEQ ID N°49.

15

La série de Figures 50 :

La série de Figures 50 représente la série de séquences nucléotidiques SEQ ID N°50.

20

Figure 51 :

A. la construction pJVED: Plasmid navette(pouvant se multiplier chez les mycobactéries ainsi que chez
25 E.coli). avec un gène de résistance à la kanamycine (issu de Tn903) comme marqueur de sélection. Le gène *phoA* tronqué(Δ *phoA*) et le gène *luc* forment un opéron synthétique.

B. Séquence de la jonction entre *phoA* et *luc*.

30

Figure 51 :

Hybridation génomique (Southern blot) de l'ADN génomique de différentes espèces mycobactériennes à
35 l'aide d'une sonde oligonucléotidique dont la séquence est la séquence comprise entre le nucléotide en

position nt 964 (extrémité 5' de la sonde) et le nucléotide en position nt 1234 (extrémité 3' de la sonde), extrémités incluses, de la séquence SEQ ID N°1.

5 Figures 53 et 54 :

Activités Luc et PhoA de *M. smegmatis* recombinant contenant le pJVED avec différents fragments nucléotidiques comme décrits en exemple. Les figures 52 et 53 représentent les résultats obtenus pour deux expériences distinctes réalisées dans les mêmes conditions.

15

Figure 55 :

Représentation de l'hydrophobicité (Kyte et Doolittle) de la séquence codante du polypeptide DP428 avec sa représentation schématique. Le motif LPISG précède immédiatement la région C-terminale hydrophobe. La séquence se termine par deux arginines.

20

Figure 56 :

25 Représentation de l'hydrophobicité (Kyte et Doolittle) de la séquence du polypeptide M1C25 de séquence d'acides aminés SEQ ID N° 30.

Figure 57 :

30

A- Gel d'acrylamide (12%) en condition dénaturante d'un extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 sans et après 4 heures d'induction par l'IPTG, coloré au bleu de Comassie.

35

ligne 1: Marqueur de masse molaire (Prestained SDS-PAGE Standards High Range BIO-RAD®).

5 ligne 2: Extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 sans induction par l'IPTG.

10 ligne 3: Extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 après 4 heures d'induction par l'IPTG.

ligne 4: Marqueur de masse molaire (Prestained SDS-PAGE Standards Low Range BIO-RAD®).

15 B- Western blot d'un gel semblable gel (acrylamide 12%) révélé grâce à l'anticorps penta-His commercialisé par la société Quiagen.

20 ligne 1: représentation du marqueur de masse molaire (Prestained SDS-PAGE Standards High Range BIO-RAD®).

25 ligne 2: extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 sans induction par l'IPTG.

30 ligne 3: extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 après 4 heures d'induction par l'IPTG.

ligne 4: représentation du marqueur de masse molaire (Prestained SDS-PAGE Standards Low Range BIO-RAD®)

35 La bande présente très majoritairement dans les lignes correspondant aux bactéries induites par l'IPTG par rapport à celles non induites par l'IPTG, comprise entre 34200 et

28400 daltons, correspond à l'expression de l'insert M1C25 cloné dans le vecteur pQE-60 (Qiagen®).

5 En ce qui concerne les légendes des autres figures qui sont numérotées par un caractère alphanumérique, chacune de ces autres figures représente la séquence nucléotidique et la séquence d'acides aminés de séquence SEQ ID dont la numérotation est identique au caractère alphanumérique de
10 chacune desdites figures.

Les numérotations alphanumériques des figures représentant les SEQ ID comportant un nombre suivi d'une lettre ont les significations suivantes :

- les numérotations alphanumériques présentant le même
15 nombre concernent une même famille de séquence rattachées à la séquence de référence SEQ ID dont la numérotation présente ce même nombre et la lettre A ;
- les lettres A, B et C pour une même famille de séquences distinguent les trois phases de lecture possibles de la
20 séquence nucléotidique SEQ ID de référence (A) ;
- les lettres indexées par un prime (') signifient que la séquence correspond à un fragment de la séquence SEQ ID de référence (A) ;
- la lettre D signifie que la séquence correspond à la
25 séquence du gène prédit par Cole et al., 1998 ;
- la lettre F signifie que la séquence correspond à la phase ouverte de lecture (ORF pour "Open Reading Frame") contenant la séquence "D" correspondante d'après Cole et al., 1998 ;
- 30 - la lettre G signifie que la séquence est une séquence prédite par Cole et al., 1998, et présentant une homologie de plus de 70% avec la séquence SEQ ID de référence (A) ;
- la lettre H signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence "G"
35 correspondante d'après Cole et al., 1998 ;
- la lettre R signifie que la séquence correspond à une séquence prédite par Cole et al., 1998, en amont de la

séquence "D" correspondante et pouvant être en phase avec la séquence "D" en raison d'erreurs de séquençage possibles ;

- la lettre P signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence "R" correspondante ;

- la lettre Q signifie que la séquence correspond à une séquence contenant les séquences "F" et "p" correspondantes.

En ce qui concerne la famille de séquences SEQ ID N° 4, l'insert précédent *phoA* contient deux fragments non contigus sur le génome, SEQ ID 4J et SEQ ID 4A, et donc issus d'un clonage multiple permettant l'expression et l'exportation de *phoA*. Ces deux fragments non contigus, les gènes et les phases ouvertes de lecture qui les contiennent d'après Cole et al., 1998, sont importants pour l'exportation d'un polypeptide antigène :

- les lettres J, K et L distinguent les trois phases de lecture possibles de la séquence nucléotidique "J" correspondante ;

- la lettre M signifie que la séquence correspond à la séquence prédite par Cole et al., 1998, et contenant la séquence SEQ ID N° 4J ;

- la lettre N signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence SEQ ID N° 4M.

En ce qui concerne la famille de séquences SEQ ID N° 45, la lettre Z signifie que la séquence correspond à la séquence d'un fragment cloné fusionné avec *phoA*.

Enfin, en ce qui concerne la famille de séquence SEQ ID N° 41, la lettre S signifie que la séquence correspond à une séquence prédite par Cole et al., 1998 et pouvant être dans la même phase de lecture que la séquence "D"

correspondante, la lettre T signifiant que la séquence correspondante contient les séquences "F" et "S" correspondantes.

EXEMPLES

Matériel et méthodes

5 Cultures bactériennes, plasmides et milieux de cultures

E. coli a été cultivé sur milieu liquide ou solide Luria-Bertani (LB). *M. smegmatis* a été cultivé sur milieu liquide Middlebrook 7H9 (Difco) additionné de dextrose albumine (ADC), 0,2 % de glycérol et 0,05 % de Tween, ou sur milieu
10 solide L. Si nécessaire, l'antibiotique kanamycine a été rajouté à une concentration de 20 $\mu\text{g/ml}^{-1}$. Les clones bactériens présentant une activité PhoA ont été détectés sur de l'agar LB contenant du 5-bromo-4-chloro-3-indolye
15 phosphate (X-P, à 40 $\mu\text{g/ml}^{-1}$).

Manipulation d'ADN et séquençage

Les manipulations d'ADN et les analyses par
20 Southern blot ont été effectuées en utilisant les techniques standard (Sambrook et al., 1989). Les séquences d'ADN double brin ont été déterminées avec un kit de séquençage Taq Dye Deoxy Terminator Cycle (Applied Biosystems), dans un Système 9600 GeneAmp PCR (Perkin-
25 Elmer), et après migration sur un système d'analyse ADN modèle 373 (Applied Biosystems).

Constructions des plasmides

30 Le plasmide pJVEDa a été construit à partir de pLA71, plasmide de transfert comportant le gène *phoA* tronqué et placé en phase avec *BlaF*. pLA71 a été coupé avec les enzymes de restriction *KpnI* et *NotI*, retirant ainsi *phoA* sans toucher le promoteur de *BlaF*. Le gène *luc* codant
35 pour la luciférase de luciole a été amplifié à partir de

pGEM-luc et un site de liaison du ribosome a été rajouté. *phoA* a été amplifié à partir de pJEM11. Les fragments amplifiés ont été coupés avec *Pst*I et ligaturés ensemble. Les oligodéoxynucléotides utilisés sont les suivants :

- 5 pPV.luc.Fw : 5'GACTGCTGCAGAAGGAGAAGATCCAAATGG3'
 luc.Bw : 5'GACTAGCGGCCGCGAATTCGTGACCTCCGAGG3'
 pJEM.phoA.Fw : 5'CCGCGGATCCGGATACGTAC3'
 phoA.Bw: 5'GACTGCTGCAGTTTATTTTCAGCCCCAGAGCG3'.

Le fragment ainsi obtenu a été réamplifié en
 10 utilisant les oligonucléotides complémentaires de ses extrémités, coupé avec *Kpn*I et *Not*I, et intégré dans pLA71 coupé avec les mêmes enzymes. La construction résultante a été électroporée dans *E. coli* DH5 α et *M. smegmatis* mc2 155. Un clone *M. smegmatis* émettant de la lumière et présentant
 15 une activité *phoA* a été sélectionné et appelé pJVED/*blaF*. L'insert a été retiré en utilisant *Bam*HI et la construction refermée sur elle-même, reconstruisant ainsi le pJVED_a. Afin d'obtenir le pJVED_{b,c}, le multisite de clonage a été coupé avec *Sca*I et *Kpn*I et refermé en enlevant un (pJVED_b)
 20 ou deux (pJVED_c) nucléotides du site *Sna*BI. Après fusion six cadres de lecture ont pu ainsi être obtenus. L'insert du pJVED/*hsp18* a été obtenu par amplification en chaîne par polymérase (ACP) de pPM1745 (Servant et al., 1995) en utilisant des oligonucléotides de la séquence :

- 25 18.Fw : 5'GTACCAGTACTGATCACCCGTCTCCCGCAC3'
 18.Back : AGTCAGGTACCTCGCGGAAGGGGTCAGTGCG3'

Le produit a été coupé avec *Kpn*I et *Sca*I, et ligaturé à pJVED_a, coupé avec les mêmes enzymes, quittant ainsi le pJVED/*hsp18*.

30

Le pJVED/P19kDa et le pJVED/*erp* furent construits en coupant avec *Bam*HI l'insert de pExp410 et pExp53 respectivement, et en les insérant dans le site *Bam*HI du

multisite de clonage de pJVED_a.

Mesure de l'activité phosphatase alcaline

La présence d'activité est détectée par la couleur
5 bleue des colonies croissant sur un milieu de culture
contenant le substrat 5-bromo 4-chloro 3-indolyl phosphate
(XP), puis l'activité peut être mesurée quantitativement de
manière plus précise de la façon suivante :

M. smegmatis ont été cultivés dans un milieu LB
10 additionnés de 0,05 % de Tween 80 (Aldrich) et de
kanamycine (20 µg/ml- 1) à 37°C pendant 24 heures.
L'activité de la phosphatase alcaline a été mesurée par la
méthode de Brockman et Heppel (Brockman et al., 1968) dans
un extrait soniqué, avec p-nitrophénylphosphate comme
15 substrat de la réaction. La quantité de protéines a été
mesurée par essai Bio-Rad. L'activité phosphatase alcaline
est exprimée en unité arbitraire (densité optique à 420 nm
x µg de protéines- 1 x minutes- 1).

20 Mesure de l'activité luciférase

M. smegmatis a été cultivé dans un milieu LB
additionné de 0,05 % de Tween 80 (Aldrich) et de kanamycine
(20 µg/ml- 1) à 37°C pendant 24 heures et utilisé en pleine
25 croissance exponentielle (DO à 600 nm comprise entre 0,3 et
0,8). Les aliquots de suspensions bactériennes ont été
brièvement soniqués et l'extrait cellulaire a été utilisé
pour mesurer l'activité de la luciférase. 25 µl de
l'extrait soniqué ont été mélangés avec 100 µl de substrat
30 (système d'essai luciférase Promega) automatiquement dans
un luminomètre et la lumière émise exprimée en ULR ou RLU
(Unités Lumineuses Relatives). Les bactéries ont été
comptées par dilutions sérielles de la suspension d'origine
sur milieu agar LB kanamycine et l'activité de la
35 luciférase exprimée en ULR/µg de protéines bactériennes ou
en ULR/10³ bactéries.

Construction de banques génomiques de *M. tuberculosis* et de *M. bovis*-BCG

5 Les banques ont été obtenues en utilisant essentiellement pJVEDa,b,c précédemment décrits.

Préparation de macrophages issus de la moelle osseuse et infection par *M. smegmatis* recombinants

10

Les macrophages issus de la moelle osseuse ont été préparés comme décrits par Lang et al., 1991. En résumé, les cellules de la moelle osseuse ont été prélevées du fémur de souris C57BL/6 agée de 6 à 12 semaines (Iffa-Credo, 15 France). Les cellules en suspensions ont été lavées et resuspendues dans du DMEM enrichi avec 10 % de sérum foetal de veau, 10 % de milieu L-cell conditionné et 2 mM de glutamine, sans antibiotiques. 10⁶ cellules ont été ensemencées sur des plaques 24 puits Costar à fond plat 20 dans 1 ml. Après quatre jours à 37°C dans une atmosphère humide à 10 % de teneur en CO₂, les macrophages ont été rincés et réincubés pendant deux à quatre jours supplémentaires. Les cellules d'un puits contrôle ont été lysées avec du triton x 100 à 0,1 % dans l'eau et les 25 noyaux énumérés. Environ 5 x 10⁵ cellules adhérentes ont été comptées. Pour l'infection, *M. smegmatis* portant les différents plasmides a été cultivé en pleine phase exponentielle (DO_{600nm} entre 0,4 et 0,8) et dilué jusqu'à une DO de 0,1 puis 10 fois dans un milieu pour macrophage. 30 1 ml a été ajouté à chaque puits et les plaques ont été centrifugées et incubées quatre heures à 37°C. Après trois lavages, les cellules ont été incubées dans un milieu contenant de l'amykacine pendant deux heures. Après trois nouveaux lavages, les cellules infectées adhérentes ont été 35 incubées dans un milieu macrophage pendant une nuit. Les cellules ont ensuite été lysées dans 0,5 ml de tampon de

lyse (Promega). 100 μ l ont été soniqués et la lumière émise a été mesurée sur 25 μ m. Simultanément, les bactéries ont été énumérées par étalement sur L-agar-kanamycine (20 μ g/ml⁻¹). La lumière émise est exprimée en ULR/10³ bactéries.

Analyses des banques de données

Les séquences nucléotidiques ont été comparées à EMBL et GenBank en utilisant l'algorithme FASTA et les séquences protéiques ont été analysées par similitude grâce aux banques de données PIR et Swiss Prot en utilisant l'algorithme BLAST .

Exemple 1 : Les vecteurs pJVED

Les vecteurs pJVED (Figure 51) sont des plasmides portant un gène *phoA* tronqué de *E. coli* dépourvu de codon d'initiation, de séquence signal et de séquence régulatrice. Le site multiple de clonage (SMC) permet l'insertion de fragments des gènes codants pour d'éventuelles protéines exportées ainsi que leurs séquences de régulation. Dès lors, la protéine de fusion peut être produite et présenter une activité phosphatase alcaline si elle est exportée. Seules les fusions en phase pourront être productives. Ainsi, le SMC a été modifié de sorte que les fusions peuvent être obtenues dans six phases de lecture. En aval de *phoA*, le gène *luc* de la luciférase de luciole a été inséré. Le gène complet avec le codon d'initiation mais sans qu'aucun promoteur n'ait été utilisé devrait ainsi s'exprimer avec *phoA* comme dans un opéron synthétique. Un nouveau site de liaison des ribosomes a été inséré huit nucléotides en amont du codon d'initiation de *luc*. Deux terminateurs transcriptionnels sont présents dans les vecteurs pJVED, un en amont du SMC et un second en aval de *luc*. Ces vecteurs sont des plasmides de transfert *E.*

coli-mycobacterium avec un gène de résistance à la kanamycine comme marqueur de sélection.

phoA et *luc* fonctionnent comme dans un opéron, mais
5 l'exportation est nécessaire pour l'activité *phoA*.

Quatre plasmides ont été construits par insertion dans le SMC de fragments d'ADN d'origine diverse :

Dans la première construction nommée pJVED/*blaF*, le fragment de 1,4 kb provient du plasmide déjà décrit pLA71
10 (Lim et al., 1995). Ce fragment issu du gène β -lactamase (*blaF*) de *M. fortuitum* D216 (Timm et al., 1994) inclut le promoteur muté hyperactif, le segment codant pour 32 acides aminés de la séquence signal et les 5 premiers acides aminés de la protéine mature. Ainsi cette construction
15 inclut le promoteur le plus fort connu chez *mycobacterium* et les éléments nécessaires à l'exportation de la protéine de la fusion *phoA*. Par conséquent, on peut attendre de cette construction une forte émission de lumière et une bonne activité *phoA* (cf figures 53 et 54).

20 Dans une deuxième construction nommée pJVED/*hsp18*, un fragment de 1,5 kb a été cloné à partir du plasmide déjà décrit pPM1745 (Servant et al., 1995). Ce fragment inclut les nucléotides codants pour les dix premiers acides aminés de la protéine de choc thermique de 18 kb issue de
25 *Streptomyces albus* (heat shock protein 18, HSP 18), le site de liaison du ribosome, le promoteur et, en amont, des sites régulateurs contrôlant son expression. Cette protéine appartient à la famille de alpha-crystalline de HSP à faible poids moléculaire (Verbon et al., 1992). Son
30 homologue issu de *M. leprae*, l'antigène de 18 kDa, est déjà connu pour être induit durant la phagocytose par un macrophage murin de la lignée cellulaire J-774 (Dellagostinet al., 1995). Dans des conditions de culture standard, le pJVED/*hsp18*, montre une faible activité *luc* et
35 aucune activité *phoA* (cf figures 53 et 54).

Dans une troisième construction, nommée pJVED/P19kDa, l'insert issu de pExp410 (Lim et al., 1995) a été coupé et cloné dans le SMC de pJVED_a. Ce fragment inclut les nucléotides codants pour les 134 premiers acides aminés de la protéine connue de *M. tuberculosis* 19 kDa et de ses séquences régulatrices. Comme cela a pu être mis en évidence, cette protéine est une lipoprotéine glycosylée (Garbe et al., 1993 ; Herrmann et al., 1996). Sur les figures 53 et 54, on observe, pour cette construction, une bonne activité luc correspondant à un promoteur fort, mais l'activité *phoA* est la plus forte des quatre constructions. L'activité *phoA* élevée de cette protéine de fusion avec une lipoprotéine s'explique par le fait qu'elle reste attachée à la paroi cellulaire par son extrémité N-terminal.

Dans la quatrième et dernière construction nommée pJVED/erp l'insert provient de pExp53 (Lim et al., 1995) et a été cloné dans le SMC de pJVED_a. pExp53 est le plasmide initial sélectionné pour son activité *phoA* et contenant une partie du gène *erp* de *M. tuberculosis* qui code pour un antigène de 28 kDa. Ce dernier inclut la séquence signal, une partie de la protéine mature et, en amont du codon d'initiation, le site de liaison de ribosome. Le promoteur a été cartographié. Une boîte fer (iron box) putative du type *fur* est présente dans cette région et encadre la région -35 du promoteur (Berthet et al., 1995). Comme prévu (figures 53 et 54) cette construction présente une bonne émission lumineuse et une bonne activité *phoA*. Le fait que cette protéine de fusion, contrairement à la fusion avec la lipoprotéine de 19 kDa, ne semble pas attachée à la paroi cellulaire n'exclut pas que la protéine native y soit associée. De plus, l'extrémité C-terminal de *erp* est absente de la protéine de fusion.

Exemple 2 : Construction d'une banque d'ADN génomique de

M. tuberculosis dans les vecteurs pJVEDs et identification d'un des membres de ces banques, (DP428), induit au cours de la phagocytose par les macrophages murins dérivés de la moelle osseuse.

5 Les différentes constructions sont testées pour leur capacité à évaluer l'expression intracellulaire des gènes identifiés par l'expression de *phoA*. Dans cet objectif, l'activité *luc* est exprimée en URL pour 10³ bactéries en culture axénique et/ou dans des conditions
10 intracellulaires. L'induction ou la répression suivant la phagocytose par les macrophages murins dérivés de la moelle osseuse peut être évaluée convenablement par la mesure des activités spécifiques. Les résultats de deux expériences distinctes sont présentés dans le tableau 2.

15 Le plasmide pJVED/*hsp18* a été utilisé comme contrôle positif pour l'induction durant la phase de croissance intracellulaire. Bien que l'induction du promoteur par le chauffage de la bactérie à 42°C n'ait pas été concluant la phagocytose de la bactérie conduit
20 clairement à une augmentation de l'activité du promoteur. Dans toutes les expériences, l'activité *luc* intracellulaire a été fortement induite, augmentant de 20 à 100 fois l'activité basale initialement faible (Servant, 1995).

Le plasmide pJVED/*blaF* a été utilisé comme contrôle
25 de la modulation non spécifique au cours de la phagocytose. De faibles variations ont pu être mises en évidence, probablement dues à des changements de conditions de cultures. Quoi qu'il en soit, ces faibles variations ne sont pas comparables à l'induction observée avec le
30 plasmide pJVED/*hsp18*.

Tous les membres de la banque d'ADN ont été testés par mesure de l'activité du promoteur durant la croissance intracellulaire. Parmi eux, le DP428 est fortement induit au cours de la phagocytose (tableaux 1 et 2).

TABLEAU 1

Construction	% Récupération	URL/10 ³ bactéries extracellulaire	URL/10 ³ bactéries intracellulaire	Induction
pJVED/blal*	0,5	1460	1727	1,2
pJVED/hsp18	0,6	8	57	7,1
pJVED/DP428	0,7	0,06	18	300
Construction	% Récupération C57BL/6 Balb/C	URL/10 ³ bactéries extracellulaire	URL/10 ³ bactéries intracellulaire C57BL/6 Balb/C	Induction C57BL/6 Balb/C
pJVED/blal*	7 1,1	662	250 911	0,4 1,4
pJVED/hsp18	6,7 1,7	164	261 325	1,6 2
pJVED/DP428	1,6 2,1	0,08	1,25 3,3	15,6 41

5 TABLEAU 2

Construction	% Récupération	URL/10 ³ bactéries extracellulaire	URL/10 ³ bactéries intracellulaire	Induction
pJVED/blal*	22	1477	367	0,25
pJVED/hsp18	7	0,26	6,8	26
pJVED/DP428	21	0,14	4	28

Le fragment nucléotidique codant pour la région N-terminale du polypeptide DP428 de séquence SEQ ID N° 28 est contenu dans le plasmide déposé à la CNCM sous le N° I-1818.

La totalité de la séquence codant pour le polypeptide DP428 a été obtenue comme détaillée ci-après.

Une sonde a été obtenue par PCR à l'aide des oligonucléotides de séquence SEQ ID N° 25 et SEQ ID N° 26. Cette sonde a été marquée par extension aléatoire en présence de ³²P dCTP. Une hybridation de l'ADN génomique de *M. tuberculosis* souche Mt103 préalablement digéré par l'endonucléase ScaI a été réalisée à l'aide de ladite

sonde. Les résultats de l'hybridation ont fait apparaître qu'un fragment d'ADN d'environ 1,7 kb était marqué. Du fait qu'il existe un site *ScaI* s'étendant du nucléotide nt 984 au nucléotide nt 989 de la séquence SEQ ID N° 1, c'est-à-dire du côté 5' de la séquence utilisée comme sonde, la fin de la séquence codante est nécessairement présente dans le fragment détecté par hybridation.

L'ADN génomique de la souche Mt 103 de *M. tuberculosis*, après digestion par *ScaI*, a subi une migration sur un gel d'agarose. Les fragments de tailles comprises entre 1,6 et 1,8 kb ont été clonés dans le vecteur pSL1180 (Pharmacia) préalablement clivé par *ScaI* et déphosphorylé. Après transformation de *E. coli* avec les vecteurs recombinants résultants, les colonies obtenues ont été criblées à l'aide de la sonde. Le criblage a permis d'isoler six colonies hybridant avec cette sonde.

Les inserts contenus dans les plasmides des clones recombinants précédemment sélectionnés ont été séquencés, puis les séquences alignées de manière à déterminer la totalité de la séquence codant pour DP428, plus spécifiquement la SEQ ID N° 2.

Un couple d'amorces a été synthétisé afin d'amplifier, à partir de l'ADN génomique de *M. tuberculosis*, souche Mt 103, la totalité de la séquence codant pour le polypeptide DP428. L'amplicon obtenu a été cloné dans un vecteur d'expression.

Des couples d'amorces appropriés pour l'amplification et le clonage de la séquence codant pour le polypeptide DP428 peuvent être aisément réalisés par l'homme du métier, sur la base des séquences nucléotidiques SEQ ID N°1 et SEQ ID N°2.

Un couple d'amorces particulier selon l'invention est le couple d'amorces suivants, capable d'amplifier l'ADN codant pour le polypeptide DP428 dépourvu de sa séquence signal :

- Amorce aller (SEQ ID N° 29), comprenant la séquence allant du nucléotide en position nt 1021 au nucléotide nt 1044 de la séquence SEQ ID N° 2 :

5' -AGTGCATGCTGCTGGCCGAACCATCAGCGAC- 3'

- Amorce retour (SEQ ID N° 30), comprenant la séquence complémentaire de la séquence allant du nucléotide en position nt 1345 au nucléotide en position nt 1325 de la séquence SEQ ID N° 2 :

5' -CAGCCAGATCTCGGGCGCCCTGCACCGCCTG- 3',

dans lesquelles la partie soulignée représente les séquences hybridant spécifiquement avec la séquence SEQ ID N° 2 et les extrémités 5' correspondent à des sites de restriction en vue du clonage de l'amplicon résultant dans un vecteur de clonage et/ou d'expression.

Un vecteur particulier utilisé pour l'expression du polypeptide DP428 est le vecteur pQE70 commercialisé par la société Qiagen.

Exemple 3 : La séquence complète du gène DP428 et de ses régions flanquantes.

Une sonde de la région codante de DP428 a été obtenue par ACP, et utilisée pour hybrider l'ADN génomique de différentes espèces de mycobactéries. D'après les résultats de la figure 3, le gène est présent uniquement dans les mycobactéries du complexe de *M. tuberculosis*.

L'analyse de la séquence suggère que DP428 pourrait faire partie d'un opéron. La séquence codante et les régions flanquantes ne présentent aucune homologie avec des séquences connues déposées dans les banques de données.

D'après la séquence codante, ce gène code pour une protéine de 10 kDa avec un peptide signal, une extrémité C-terminal hydrophobe terminée par deux arginines et précédée par un motif LPISG semblable au motif connu LPXTG. Ces deux

arginines pourraient correspondre à un signal de rétention et la protéine DP428 pourrait être accrochée par ce motif à des peptidoglycanes comme cela a déjà été décrit chez d'autres bactéries Gram⁺ (Navarre et al., 1994 et 1996).

5 Le mécanisme de survie et de croissance intracellulaire des mycobactéries est complexe et les relations intimes entre la bactérie et la cellule hôte restent inexplicées. Quel que soit le mécanisme, la croissance et la survie intracellulaire des mycobactéries
10 dépend de facteurs produits par la bactérie et capables de moduler la réponse de l'hôte. Ces facteurs peuvent être des molécules exposées à la surface cellulaire telle que LAM ou des protéines associées à la surface cellulaire, ou des molécules activement secrétées.

15 D'un autre côté, intracellulairement, les bactéries elles-mêmes doivent faire face à un environnement hostile. Elles semblent y répondre par des moyens proches de ceux mis en oeuvre dans les conditions de stress, par l'induction de protéines de choc thermique (Dellagostin et
20 al., 1995), mais aussi par induction ou la répression de différentes protéines (Lee et al., 1995). En utilisant une méthodologie dérivée de la PCR, Plum et Clark-curtiss (Plum et al., 1994) ont montré qu'un gène de *M. avium* inclu dans un fragment d'ADN de 3 kb, est induit après la phagocytose
25 par des macrophages humains. Ce gène code pour une protéine exportée comprenant une séquence leader mais ne présentant pas d'homologie significative avec les séquences proposées par les banques de données. L'induction, pendant la phase de croissance intracellulaire, d'une protéine de choc
30 thermique de faible poids moléculaire issue de *M. leprae* a également été mise en évidence (Dellagostin et al., 1995). Dans une autre étude, les protéines bactériennes de *M. tuberculosis* ont été métaboliquement marquées pendant la phase de croissance intracellulaire ou bien dans des
35 conditions de stress et séparées par électrophorèse sur gel à deux dimensions : 16 protéines de *M. tuberculosis* ont été

induites et 28 réprimées. Les mêmes protéines sont mises en jeu au cours de stress provoqué par un faible pH, un choc thermique, H₂O₂, ou au cours de la phagocytose par des monocytes humains de la lignée THP1. Quoi qu'il en soit, le comportement des protéines induites et réprimées était unique dans chaque condition (Lee et al., 1995). Pris ensemble, ces résultats indiquent qu'un dialogue moléculaire subtile est mis en place entre les bactéries et leurs hôtes cellulaires. De ce dialogue dépend probablement le sort de l'organisme intracellulaire.

Dans ce contexte, l'induction de l'expression de DP428 pourrait être d'une importance majeure, indiquant un rôle important de cette protéine dans la survie et la croissance intracellulaire.

La méthode utilisée dans ces expériences pour évaluer l'expression intracellulaire des gènes (cf. Jacobs et al., 1993, pour la méthode de détermination de l'expression de la luciférase de luciole, et Lim et al., 1995, pour la méthode de détermination de l'expression du gène *PhoA*) présente l'avantage d'être simple comparée aux autres techniques comme la technique décrite par Mahan et al. (Mahan et al., 1993) adaptée aux mycobactéries et proposée par Bange et al. (Bange et al., 1996), ou la méthode substractive basée sur l'ACP décrite par Plum et Clark-curtiss (Plum et al., 1994). Il existe indiscutablement une variabilité comme le montre la comparaison des différentes expériences. Bien que provoquer l'induction ou la répression soit suffisant, il est désormais possible de l'évaluer fournissant ainsi un outil supplémentaire d'études physiologiques des protéines exportées identifiées par fusion avec *phoA*.

Exemple 4 :

Recherche d'une modulation de l'activité des promoteurs lors des phases intramacrophagiques.

Des macrophages de moelle osseuse de souris sont préparés comme décrit par Lang et Antoine (Lang et al., 1991). Les bactéries de *M. segmentis* recombinantes, dont on a déterminé l'activité luciférase par 10^3 bactéries comme précédemment, sont incubées à 37°C sous atmosphère humidifiée et enrichie en CO₂ à 5%, pendant 4 heures en présence de ces macrophages de telle manière qu'elles soient phagocytées. Après rinçage pour éliminer les bactéries extracellulaires restantes, on ajoute au milieu de culture de l'amikacine (100 µg/ml) pendant deux heures. Après un nouveau rinçage, le milieu est remplacé par un milieu de culture (DMEM enrichi de 10 % de sérum de veau et 2 mM de glutamine) sans antibiotiques. Après une nuit d'incubation comme précédemment, les macrophages sont lysés à froid (4°C) à l'aide d'un tampon de lyse (cee lysis buffer, Promega), et l'activité luciférase par 10^3 bactéries déterminée. Le rapport des activités à la mise en culture et après une nuit donne le coefficient d'induction.

Exemple 5 :

Isolement d'une série de séquences par séquençage directement à partir des colonies.

Une série de séquences permettant l'expression et l'exportation de *phoA* ont été isolées à partir de l'ADN de *M. Tuberculosis* ou de *M. Bovis* BCG. Parmi ce groupe de séquences, deux d'entre elles ont été d'avantage étudiées, les gènes entiers correspondant aux inserts ont été clonés, séquencés, et des anticorps contre le produit de ces gènes ont servi à montrer en microscopie électronique que ces gènes codaient pour des antigènes retrouvés à la surface des bacilles de la tuberculose. L'un de ces gènes encode pour une séquence signal d'exportation consensus, l'autre ne possédait aucune caractéristique de gène codant pour une protéine exportée, d'après la séquence. Un

autre gène DP428 a été séquencé avant que la séquence du génome de *M. Tuberculosis* ne soit disponible. Il contient une séquence ressemblant à la séquence consensus d'attachement au peptidoglycane, ce qui suggère qu'il s'agit aussi d'un antigène vraisemblablement retrouvé à la surface des bacilles de la tuberculose. L'étude des trois gènes *erp*, *des*, et celui codant pour DP428 montre que le système *phoA* que nous avons développé chez les mycobactéries permet de repérer des gènes codant pour des protéine exportées sans déterminant repérable par des études *in silico*. Ceci est particulièrement vrai pour les polypeptides qui ne possèdent pas de séquence signal consensus (*des*) ou non pas de similarité avec des protéines de fonction connue (*erp* et DP428).

Un certain nombre d'inserts ont été identifiés et séquencés avant la connaissance du génome de *M. Tuberculosis*, d'autres après. Ces séquences peuvent être considérées comme des amorces permettant de rechercher des gènes codant pour des protéines exportées. A ce jour, une série d'amorces ont été séquencées et les gènes entiers correspondants ont été soit séquencés, soit identifiés d'après la séquence publiée du génome. Pour tenir compte des erreurs de séquençage toujours possibles, les régions en amont ou en aval de certaines amorces ont été considérées comme pouvant faire partie de séquences codant pour des protéines exportées. Dans certains cas des similarités avec des gènes codant pour des protéines exportées ou des séquences caractéristiques de signaux d'exportation ou des caractéristiques topologiques de protéines membranaires ont été détectées.

Des séquences amorces s'avèrent correspondre à des gènes appartenant à des familles de gènes possédant plus de 50 % de similarité. On peut ainsi indiquer que les autres gènes détectés par similarité avec une amorce codent pour

des protéines exportées. C'est le cas de la séquence SEQ ID N° 8G et SEQ ID N° 8H possédant plus de 77 % de similarité avec SEQ ID N° 8A'.

Les séquences pouvant coder pour des protéines exportées sont les suivantes : SEQ ID N° 1, 8, 9, 8G, 8H, 13, 3, 10, 19, 20, 6, 16, 22, 23, 24, 39, 44, 46, et 50.

Des gènes identifiés d'après les amorces à partir de la séquence du génome n'ont aucune caractéristique (d'après la séquence) de protéines exportées. Il s'agit des séquences suivantes : SEQ ID N° 4, 27, 11, 12, 14, 7, 15, 17, 18, 21, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 45, 47, 48, et 49.

D'après la séquence d'autres organismes comme *E. coli*, on peut rechercher dans la séquence du génome de *M. tuberculosis*, des gènes possédant des similarités avec des protéines connues pour être exportées chez d'autres organismes bien que ne possédant pas de séquence signal d'exportation. Dans ce cas une fusion avec *phoA* est un protocole avantageux pour déterminer si ces séquences de *M. tuberculosis* codent pour des protéines exportées bien que ne présentant pas de séquence signal consensus. Il a été en effet possible de cloner SEQ ID N° 49, une séquence similaire à un gène de *E. coli* de la famille *htrA*. Une fusion de SEQ ID N° 49 avec *phoA* conduit à l'expression et à l'exportation de *phoA*. Des colonies *M. smegmatis* hébergeant une fusion SEQ ID N° 49 *phoA* sur un plasmide pJVED sont bleues.

SEQ ID N° 49 est donc considérée comme une protéine exportée.

La méthode *phoA* est donc utile pour détecter d'après la séquence de *M. Tuberculosis* des gènes codant pour des protéines exportées sans qu'ils ne codent pour des séquences caractéristiques des protéines exportées.

Même si une séquence possède des déterminants de protéines exportées, cela ne démontre pas une exportation fonctionnelle. Le système *phoA* permet de montrer que le gène suspecté code réellement pour une protéine exportée.

- 5 Ainsi, il a été vérifié que la séquence SEQ ID N° 50 possédait bien des signaux d'exportation.

TABEAU 3

SEQ ID N°	Référence de la séquence correspondante prédite par Cole et al.		Annotation
SEQ ID N°1	Rv 0203	*	Séquence hydrophobe en N-terminal
SEQ ID N°4			
SEQ ID N°27	Rv 2050		Pas de prédiction
SEQ ID N°8			
SEQ ID N°9	Rv 2563	*	Protéine membranaire
SEQ ID N°			
BG', H'	Rv 0072	*	Possible protéine de transport transmembranaire de type ABC
SEQ ID N°11	Rv 0546c	ML	Protein S-D Lactoyl Glutathione-méthyl glyoxal lyase
SEQ ID N°12	pas de prédiction		non retrouvé dans <i>M. tuberculosis</i> H37rv
SEQ ID N°13			
SEQ ID N°3	Rv 1984c	*	probable précurseur cutinase avec une séquence signal N-terminale
SEQ ID N°10			
SEQ ID N°14			
SEQ ID N°7	pas de prédiction		pas de prédiction
SEQ ID N°15	avec décalage de lecture, pourrait être en phase avec Rv 2530c		pas de prédiction
SEQ ID N°17	Rv 1303	ML	pas de prédiction
SEQ ID N°18	Rv 0199	ML	pas de prédiction
SEQ ID N°19	Rv 0418	*	site de fixation de lipo-protéine membranaire procaryote, similarité avec la N-acétyl puromycyne acétyl hydrolase
SEQ ID N°20	Rv 3576	*	contient un site de fixation de lipoprotéine membranaire procaryote, similarité avec une
SEQ ID N°6			

			sérine/thréonine protéine kinase
SEQ ID N° 21	Rv 3365c	ML	similarité avec une métallo peptidase à zinc
SEQ ID N° 31	non prédite		pas de prédiction
SEQ ID N° 32	Rv 0822c	ML	Existence d'une région consensus avec la famille drac
SEQ ID N° 33	Rv 1044		pas de prédiction
SEQ ID N° 34	non prédite		pas de prédiction
SEQ ID N° 35	Rv 2169c		pas de prédiction
SEQ ID N° 36	Rv 3909	ML	pas de prédiction
SEQ ID N° 37	Rv 2753c		similarité avec des dihydropicolinate synthases
SEQ ID N° 38	Rv 0175		pas de prédiction
SEQ ID N° 39	Rv 3006	ML	prédiction de séquence signal de lipoprotéine
SEQ ID N° 40	Rv 0549c		pas de prédiction
SEQ ID N° 41	Rv 2975c pouvant être en phase avec Rv 2974c		similarité avec protéine de substilis
SEQ ID N° 42	Rv 2622		similarité avec une méthyl transférase
SEQ ID N° 43	Rv 3278c	ML	pas de prédiction
SEQ ID N° 44	Rv 0309	*	pas de prédiction
SEQ ID N° 45	Rv 2169c	ML	pas de prédiction
SEQ ID N° 46	Rv 1411c	*	probable lipoprotéine avec une séquence signal N-terminale
SEQ ID N° 47	Rv 1714		similarité avec une gluconate 3-déhydrogénase
SEQ ID N° 48	Rv 0331		similarité avec une sulfide déhydrogénase et une sulfide quinone réductase
SEQ ID N° 49	Rv 0983	ML	Similarité avec une sérine protéase HtrA

SEQ ID N°5			
SEQ ID N°16	Rv 3810	* ML	Protéine de surface Berthelet et al. 1995
SEQ ID N°22 SEQ ID N°23 SEQ ID N°24	Rv 3763	*	Contient un site de fixation de lipoprotéine membranaire eucaryote
SEQ ID N°50	Rv 0125	*	Site actif des sérines protéases Séquence signal N-terminale possible

Légende du tableau 3 :

Correspondance des séquences selon l'invention avec les
séquences prédites par Cole et al. 1998, Nature, 393, 537-
544.

* : Prédiction que la protéine codée par la séquence
soit exportée

ML : Prédiction de similarité avec *M. leprae*.

Exemple 6 :

Caractéristiques et obtention de la protéine M1C25

L'extrémité N terminale de la protéine M1C25 a été
détectée par le système *PhoA* comme permettant l'exportation
de la protéine de fusion, nécessaire à l'obtention de son
activité phosphatase.

La séquence d'ADN codant pour l'extrémité N terminale de
la protéine M1C25 est contenue dans la séquence SEQ ID N°
20 de la présente demande de brevet.

A partir de cette séquence amorce, le gène complet
codant pour la protéine M1C25 a été recherché dans le
génome de *M. tuberculosis* (Fondation Wellcome Trust, site
Sanger).

Le centre Sanger a attribué à M1C25 les noms:

Rv3576,
MTCY06G11.23,
pknM

- 5 Séquence SEQ ID N° 29 du gène complet M1C25 (714 bases):
cf. Figure 29

Ce gène code pour une protéine de 237 AA, de 25 kDa de
masse molaire. Cette protéine est référencée dans les
10 banques sous les appellations:
PID:e306716,
SPTREMBL:P96858

- 15 Séquence SEQ ID N° 30 de la protéine M1C25 (237 acides
aminés): cf. Figure 30

M1C25 contient un site de fixation à la partie
lipidique des lipoprotéines de membrane des procaryotes
(PS00013 Prokaryotic membrane lipoprotein lipid attachment
20 site:
CTGGTCGGTG CGTG CATGCT CGCAGCCGGA TGC).

La fonction de M1C25 n'est pas certaine mais elle
possède très probablement une activité "sérine/thréonine-
protéine kinase". Des ressemblances sont à noter avec la
25 moitié C terminale de K08G_MYCTU Q11053 Rv1266c
(MTCY50.16). Des similarités sont aussi retrouvées avec
KY28_MYCTU.

En 5' du gène codant pour M1C25 se trouve un gène
codant potentiellement pour une protéine régulatrice
30 (PID:e306715, SPTREMBL:P96857, Rv3575c, (MTCY06G11.22c))

Le profil d'hydrophobicité (Kyte et Doolittle) de M1C25
est représenté à la figure 56.

Un site de clivage de la séquence signal est prédit
(SignalP V1.1; World Wide Web Prediction Server, Center for
35 Biological Sequence Analysis) entre les acides aminés 31 et
32: AVA-AD. Ce site de coupure est derrière un motif "AXA"
classique. Cette prédiction est compatible avec le profil

d'hydrophobicité. Dans cette séquence signal potentielle il est a remarqué la répétition trois fois de la séquence des trois acides aminés LAA.

- 5 Clonage du gène M1C25 en vue de la production de la protéine qu'il code:

Un couple d'amorces a été synthétisé afin d'amplifier, à partir de l'ADN génomique de *M. tuberculosis*, souche
10 H37Rv, la totalité de la séquence codant pour le polypeptide M1C25. L'amplicon obtenu a été cloné dans un vecteur d'expression.

Des couples d'amorces appropriés pour l'amplification
15 et le clonage de la séquence codant pour M1C25 ont été synthétisés :

-amorce aller :

5' -ATAATACCATTGGGCAAGCAGCTAGCCGCGC- 3'

-amorce retour :

20 5' -ATTTATAGATCTCTGCTTAGCAACCTTGGCCGCG- 3'

La partie soulignée représente les séquences hybridant spécifiquement avec la séquence M1C25 et les extrémités 5' correspondent à des sites de restriction en vue du clonage
25 de l'amplicon résultant dans un vecteur de clonage et/ou d'expression.

Un vecteur particulier utilisé pour l'expression du polypeptide M1C25 est le vecteur pQE60 commercialisé par la société Qiagen, en suivant le protocole et les
30 recommandations proposés par cette marque.

Les cellules utilisées pour le clonage sont des bactéries : *E. coli* XL1-Blue (résistante à la tétracycline).

Les cellules utilisées pour l'expression sont des
35 bactéries : *E. coli* M15 (résistante à la kanamycine) contenant le plasmide pRep4 (M15 pRep4).

La production de la protéine MC25 est illustrée par les figures 57 A et B. (Extraits bactériens de la souche E. coli M15 contenant le plasmide pM1C25. Les cultures bactériennes et les extraits sont préparés selon Sambrook et al. (1989). L'analyse des extraits bactériens est effectuée selon les instructions de Quiagen (1997).

Références bibliographiques

- AIDS therapies, 1993, in *Mycobacterial infections*, ISBN 0-9631698-1-5, pp. 1-11.
- 5 Altschul, S.F. et al., 1990, *J. Mol. Biol.*, 215 : 403-410.
- Andersen, P. et al., 1991, *Infect. Immun.*, 59 :1905-1910.
- Andersen, P. et al., 1995, *J. Immunol.*, 154, 3359-3372.
- Bange, F.C., A.M. Brown, and W.R. Jacobs JR., 1996, Leucine auxotrophy restricts growth of *Mycobacterium bovis* BCG in
- 10 macrophages. *Infect. Immun.*, 64,: 1794-1799.
- Barany, F., 1911, *Proc. Natl. Acad. Sci. USA*, 88 :189-193.
- Bates, J. et al., 1986, *Am. Rev. Respir. Dis.*, 134 :415-417.
- Bates, J. 1979. *Chest*. 76(Suppl.):757-763.
- 15 Bates, J. et al.. 1986. *Am. Rev. Respir. Dis.* 134 :415-417.
- Berthet, F.X., J. Rauzier, E.M. Lim, W. Philipp, B. Gicquel, and D. Portnoï, 1995, Characterization of the *M. tuberculosis* *erp* gene encoding a potential cell surface protein with repetitive structures. *Microbiology*. In press
- 20 Borremans, M. et al., 1989, *Biochemistry*, 7 : 3123-3130.
- Bouvet, E. 1994. *Rev. Fr. Lab.* 273 :53-56.
- Brockman, R.W. and Heppel L.A., 1968, On the localization of alkaline phosphatase and cyclic phosphodiesterase in *Escherichia coli*, *Biochemistry*, 7 : 2554-2561.
- 25 Burg, J.L. et al., 1996, *Mol. and Cell. Probes*, 10 :257-271.
- Chevrier, D. et al., 1993, *Mol. and Cell. Probes*, 7 :187-197.
- Clemens, D.L., 1996, Characterization of the *Mycobacterium tuberculosis* phagosome, *Trends Microbiol.*, 4 : 113-118.
- 30 Chu ,B.C.F. et al., 1986, *Nucleic Acids Res.*, 14 :5591-5603.

- Clemens, D.L. and Horwitz M.A., 1995, Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited, J. Exp. Med., 181 : 257-270.
- 5 Colignon J.E., 1996. Immunologic studies in humans. Measurement of proliferative responses of cultured lymphocytes. Current Protocols in Immunology, NIH, 2, Section II.
- Daniel, T.M. et al. 1987. Am. Rev. Respir. Dis., 135 :1137-10 1151).
- Dellagostin, O.A., Esposito G., Eales L.-J., Dale J.W. and. McFadden J.J., 1995, Activity of mycobacterial promoters during intracellular and extracellular growth. Microbiol., 141 : 2123-2130.
- 15 Drake, T.A. et al. 1987. J. Clin. Microbiol. 25:1442-1445.
- Dramsai et al., 1997, Infection and Immunity, 65, 5 : 1615-1625.
- Duck, P. et al., 1990, Biotechniques, 9:142-147.
- Erlich, H.A. 1989. In PCR Technology. Principles and 20 Applications for DNA Amplification. New York: Stockton Press.
- Felgner et al., 1987, Proc. Natl. Acad. Sci., 84:7413.
- Fraley et al., 1980, J. Biol. Chem., 255:10431.
- Gaillard, J.L., Berche P., Frehel C., Gouin E. and Cossart 25 P., 1991, Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from Gram-positive cocci, Cell., 65 : 1127-1141.
- Garbe, T., Harris D., Vordermeir M., Lathigra R., Ivanyi J. and Young D., 1993, Expression of the *Mycobacterium* 30 *tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation, Infect. Immun., 61 : 260-267.

- Guatelli, J.C. et al., 1990, Proc. Natl. Acad. Sci. USA, 87:1874-1878.
- Harboe et al., 1996, Infect. Immun., 64, 16-22.
- Herrmann, J.L., O'Gaora P., Gallagher A., Thole J.E.R. and
5 Young D.B., 1996, Bacterial glycoproteins: a link between
glycosylation and proteolytic cleavage of a 19 kDa antigen
from *Mycobacterium tuberculosis*, EMBO J. 15 : 3547-3554.
- Houbenweyl, 1974, in Meuthode der Organischen Chemie, E.
Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart.
- 10 Huygen, K. et al., 1996, Nature Medicine, 2(8):893-898.
- Innis, M.A. et al. 1990. in PCR Protocols. A guide to
Methods and Applications. San Diego: Academic Press.
- Isberg, R.R., Voorhis D.L. and Falkow S., 1987,
Identification of invasins: a protein that allows enteric
15 bacteria to penetrate cultured mammalian cells, Cell, 50 :
769-778.
- Jacobs, W.R. et al., 1991. Construction of mycobacterial
genomic libraries in shuttle cosmids. Genetic Systems for
Mycobacteria, Methods in Enzymology, 204 : 537-555.
- 20 Jacobs, W.R. et al., 1993, Science, 260 :819-822.
- Kaneda, et al., 1989, Science, 243:375
- Kiehn, T.E., et al. 1987. J. Clin. Microbiol. 25 :1551-
1552.
- Kievitis ,T. et al., 1991, J. Virol. Methods, 35 :273-286.
- 25 Kohler, G. et al., 1975, Nature, 256(5517):495-497.
- Kwob, D.Y. et al., 1989, Proc. Natl. Acad. Sci. USA, 86
:1173-1177.
- Landegren ,U. et al., 1988, Science, 241,:1077-1080.
- Lang, T. and Antoine J.-C., 1991, Localization of MHC
30 classII molecules in murine bone marrow-derived
macrophages. Immunology, 72 : 199-205.

- Lee, B.Y. and Horwitz M.A., 1995, Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*, J. Clin. Invest., 96 : 245-249.
- Lim, E.M., Rauzier J., Timm J., Torrea G., Murray A.,
5 Gicquel B. and Portnoi D., 1995, Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins, using *phoA* gene fusions, J. Bacteriol., 177 : 59-65.
- Lizardi, P.M. et al., 1988, Bio/technology, 6 :1197-1202.
- 10 Mahan, M.J. et al., 1993. Selection of bacterial virulence genes that are specifically induced in host tissues, Science, 259 : 686-688.
- Manoil L., Mekolanos J.J. and Beckwith J., J. Bacteriol., 1990, 172, 515-518.
- 15 Matthews, J.A. et al., 1988, Anal. Biochem., 169:1-25.
- Merrifield, R.D., 1966, J. Am. Chem. Soc., 88(21):5051-5052.
- Midoux, 1993, Nucleic Acids Research, 21:871-878/
- Miele, E.A. et al., 1983, J. Mol. Biol., 171:281-295.
- 20 Minton, N.P., 1984, Gene, 31, 269-273.
- Montgomery et al., 1993, DNA Cell Biol., 12:777-783.
- Navarre, W.W. et al., 1994, Molecular Microbiologie, 14(1):115-121.
- Navarre, W.W. et al., 1996, J. of Bacteriology, 178, 2 :441-
25 446.
- Pagano et al., 1967, J. Virol., 1 :891
- Pastore, 1994, Circulation, 90:I-517.
- Patel, et al. 1990. J. Clin. Microbiol. :513-518.
- Prentki, B. et Krish H.M., 1984, Gene, 29 : 303-313.
- 30 Pettersson R., Nordfelth J., Dubinina E., Bergman T., Gustafsson M., Magnusson K.E. and Wolf-Watz H., 1996,

- Modulation of virulence factor expression by pathogen target cell contact. *Science.*, 273 : 1231-1233.
- Plum, G. and Clark-Curtiss J.E., 1994, Induction of *Mycobacterium avium* gene expression following phagocytosis
- 5 by human macrophages. *Infect. Immun.*, 62 : 476-483.
- Roberts, M.C., et al. 1987. *J. Clin. Microbiol.* 25,:1239-1243.
- Rolfs, A. et al. 1991. In PCR Topics. Usage of Polymerase Chain reaction in Genetic and Infectious Disease. Berlin:
- 10 Springer-Verlag.
- Sambrook, J. et al. 1989. In Molecular cloning : A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sanchez-Pescador, R., 1988, *J. Clin. Microbiol.*,
- 15 26(10),:1934-1938.
- Schneewind, O. et al., 1995, *Science*, 268 : 103-106.
- Segev D., 1992, in « Non-radioactive Labeling and Detection of Biomolecules ». Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- 20 Servant, P. and Mazodier P., 1995, Characterization of *Streptomyces albus* 18-kilodalton heat shock-responsive protein. *J. Bacteriol.*, 177 : 2998-3003.
- Shiver, J.W., 1995, in *Vaccines 1995*, eds Chanock, R.M. Brown, F. Ginsberg, H.S. & Norrby, E.), pp.95-98, Cold
- 25 Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sorensen et al., 1995, *Infect. Immun.*, 63, 1710-1717.
- Stone, B.B. et al., 1996, *Mol. and Cell. Probes*, 10 :359-370.
- 30 Stover, C.K., Bansal G.P., Hanson M.S., Burlein S.R., Palaszynski S.R., Young J.F., Koenig S., Young D.B., Sadziene A. and Barbour A.G., 1993, Protective immunity elicited by recombinant Bacille Calmette-Guerin (BCG)

- expressing outer surface protein A (OspA) lipoprotein: a candidate Lyme disease vaccine. J. Exp. Med., 178 : 197-209.
- Sturgill-Koszycki, S., Schlesinger P.H., Chakroborty P.,
5 Haddix P.L., Collins H.L., Fok A.K., Allen R.D., Gluck S.L., Heuser J. and Russell D.G., 1994, Lack of acidification in *Mycobacterium* phagosomes by exclusion of the vesicular proton-ATPase. Science., 263 : 678-681.
- Tascon, R.E et al., 1996, Nature Medicine, 2(8):888-892.
- 10 Technique assemblage oligonucléotides, 1983, Proc. Natl. Acad. Sci. USA, 80 :7461-7465.
- Technique des bêta-cyanethylphosphoramidites, 1986, Bioorganic Chem., 4 :274-325.
- Thierry, D. et al., 1990, Nucl. Acid Res., 18 :188.
- 15 Timm, J., Perilli M.G., Duez C., Trias J., Orefici G., Fattorini L., Amicosante G., Oratore A., Boris B., Frere J.M., Pugsley A.P. and Gicquel B., 1994, Transcription and expression analysis, using *lacZ* and *phoA* gene fusions, of *Mycobacterium fortuitum* B-lactamase genes cloned from a
20 natural isolate and a high-level B-lactamase producer. Mol. Microbiol., 12 : 491-504.
- Tuberculosis Prevention Trial, 1980, Mendis, « Trial of BCG vaccines in South India for Tuberculosis Infection », Indian Journal of Medical research, 1972 (Suppl.):1-74.
- 25 Urdea, M.S. et al., 1991, Nucleic Acids Symp. Ser., 24 :197-200.
- Urdea, M.S., 1988, Nucleic Acids Research, 11: 4937-4957.
- Verbon, A., Hartskeerl R.A., Schuitema A., Kolk A.H., Young D.B. and Lathigra R., 1992, The 14,000-molecular-weight
30 antigen of *Mycobacterium tuberculosis* is related to the alpha-crystallin family of low-molecular-weight heat shock proteins. J Bacteriol., 174 : 1352-1359.

- Walker, G.T. et al., 1992, Nucleic Acids Res., 20:1691-1696.
- Walker, G.T. et al., 1992, Proc. Natl. Acad. Sci. USA, 89:392-396.
- 5 Wiker, H.G. et al., 1992, Microbiol. Rev., 56 :648-661.
- Yamaguchi, R. et al., 1989, Infect. Immun., 57 :283-288 ;
- Xu, S., Cooper A., Sturgill-Koszycki S., van Heyningen T., Chatterjee D., Orme I., Allen P. and Russel D.G., 1994, Intracellular trafficking in *Mycobacterium tuberculosis* and
- 10 *Mycobacterium avium*-infected macrophages, J. Immuno., 153: 2568-2578.
- Young, D.B. et al., 1992, Mol. Microbiol., 6 :133-145.
- Yuen, L.K.W. et al., 1993, J. Clin. Microbiol., 31 : 1615-1618.

REVENDICATIONS

1. Vecteur recombinant de criblage, de clonage et/ou d'expression, caractérisé en ce qu'il se réplique chez des mycobactéries et en ce qu'il contient :
 - 1) un réplicon fonctionnel chez les mycobactéries ;
 - 2) un marqueur de sélection ;
 - 3) une cassette rapporteur comprenant :
 - a) un site de clonage multiple (polylinker),
 - b) éventuellement un terminateur de transcription actif chez les mycobactéries, en amont du polylinker,
 - c) une séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine, ladite séquence nucléotidique étant dépourvue de son codon d'initiation et de ses séquences de régulation, et
 - d) une séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de promoteurs contenus dans le même fragment, ladite séquence nucléotidique étant pourvue de son codon d'initiation.
2. Vecteur recombinant selon la revendication 1, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine est une séquence codante issue du gène *phoA* de la phosphatase alcaline.
3. Vecteur recombinant selon l'une des revendications 1 et 2, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine est une séquence codante du gène de la β -agarase, de la nucléase d'un staphylocoque ou de la β -lactamase d'une mycobactérie.

4. Vecteur recombinant selon l'une des revendications 1 à 3, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de promoteurs contenus dans le même fragment est une séquence
5 codante issue du gène luc de la luciférase de luciole.
5. Vecteur recombinant selon l'une des revendications 1 à 4, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de
10 promoteurs contenus dans le même fragment est une séquence codante issue du gène GFP de la Green Fluorescent Protein.
6. Vecteur recombinant selon l'une des revendications 1 à 5, caractérisé en ce que le terminateur de transcription
15 actif chez les mycobactéries est le terminateur du coliphage T4 (tT4).
7. Vecteur recombinant selon l'une des revendications 1 à 6, caractérisé en ce qu'il est un plasmide choisi parmi les
20 plasmides suivants déposés à la CNCM (Collection Nationale de Cultures de Microorganismes, Paris, France) :
- a) pJVEDa déposé à la CNCM sous le N° I-1797, le 12/12/1996,
- b) pJVEDb déposé à la CNCM sous le N° I-1906, le 25 juillet
25 1997,
- c) pJVEDc déposé à la CNCM sous le N° I-1799, le 12/12/1996.
8. Vecteur recombinant selon l'une des revendications 1 à
30 7, caractérisé en ce qu'il comprend en l'un des sites de clonage du polylinker une séquence d'acide nucléique de mycobactérie chez laquelle on détecte un polypeptide susceptible d'être exporté et/ou sécrété, et/ou d'être induit ou réprimé lors de l'infection par ladite
35 mycobactérie ou encore exprimé ou produit de façon constitutive, ainsi que les séquences promotrices et/ou régulatrices associées susceptibles de permettre ou de

favoriser l'exportation et/ou la sécrétion dudit polypeptide, ou tout ou partie de gène codant pour ledit polypeptide.

- 5 9. Vecteur recombinant selon l'une des revendications 1 à 8, caractérisé en ce que la séquence d'acide nucléique de mycobactérie qu'il contient est obtenue par fragmentation physique ou par digestion enzymatique de l'ADN génomique ou de l'ADN complémentaire d'un ARN d'une mycobactérie.

10

10. Vecteur recombinant selon l'une des revendications 1 à 9, caractérisé en ce que ladite mycobactérie est *M. tuberculosis*.

- 15 11. Vecteur recombinant selon l'une des revendications 1 à 9, caractérisé en ce que ladite mycobactérie est choisie parmi *M. africanum*, *M. bovis*, *M. avium* ou *M. leprae*.

- 20 12. Vecteur recombinant selon la revendication 10, caractérisé en ce qu'il est un plasmide choisi parmi les plasmides suivants déposés à la CNCM :

- a) p6D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1814,
b) p5A3 déposé le 28 janvier 1997 à la CNCM sous le N°I-1815,
25 c) p5F6 déposé le 28 janvier 1997 à la CNCM sous le N°I-1816,
d) p2A29 déposé le 28 janvier 1997 à la CNCM sous le N°I-1817,
30 e) pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818,
f) p5B5 déposé le 28 janvier 1997 à la CNCM sous le N°I-1819,
g) p1C7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1820,
35 h) p2D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1821,

- i) p1B7 déposé le 31 janvier 1997 à la CNCM sous le N°I-1843,
j) pJVED/*M. tuberculosis* déposé le 25 juillet 1997 à la CNCM sous le N° I-1907,
5 k) pM1C25 déposé le 4 août 1998 à la CNCM sous le n°I-2062.

13. Vecteur recombinant selon la revendication 12, caractérisé en ce qu'il s'agit du plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818.

10

14. Procédé de criblage de séquences de nucléotides issues de mycobactéries pour déterminer la présence de séquences correspondant à des polypeptides exportés et/ou sécrétés pouvant être induits ou réprimés lors de l'infection, leurs
15 séquences promotrices et/ou régulatrices associées susceptibles notamment de permettre ou de favoriser l'exportation et/ou la sécrétion desdits polypeptides d'intérêt, ou tout ou partie de gènes d'intérêt codant pour lesdits polypeptides, caractérisé en ce qu'il met en oeuvre
20 un vecteurselon l'une des revendications 1 à 13.

15. Procédé de criblage selon la revendication 14, caractérisé en ce qu'il comprend les étapes suivantes :

- a) la fragmentation physique des séquences d'ADN de
25 mycobactéries ou leur digestion par au moins une enzyme déterminée et la récupération des fragments obtenus ;
b) l'insertion des fragments obtenus à l'étape a) dans un site de clonage, compatible le cas échéant avec l'enzyme de l'étape a), du polylinker d'un vecteur selon l'une des
30 revendications 1 à 13 ;
c) si besoin, l'amplification desdits fragments contenus dans le vecteur, par exemple par répllication de ce dernier après insertion du vecteur ainsi modifié dans une cellule déterminée, de préférence *E coli* ;
35 d) la transformation de cellules hôtes par le vecteur amplifié à l'étape c), ou en l'absence d'amplification, par le vecteur de l'étape b) ;

- e) la culture de cellules hôtes transformées dans un milieu permettant la mise en évidence du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs contenu dans le vecteur ;
- 5 f) la détection des cellules hôtes positives (colonies positives) pour l'expression du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs ;
- 10 g) l'isolement de l'ADN des colonies positives et l'insertion de cet ADN dans une cellule identique à celle de l'étape c) ;
- 15 h) la sélection des insertions contenues dans le vecteur, permettant l'obtention de clones positifs pour le marqueur d'exportation et/ou de sécrétion, et /ou pour le marqueur d'activité de promoteurs ;
- i) l'isolement et la caractérisation des fragments d'ADN de mycobactéries contenues dans ces insérats, et l'étape i) du procédé pouvant comporter en outre une étape de séquençage des insertions sélectionnées.
- 20
16. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie, caractérisée en ce qu'elle est obtenue par un procédé selon la revendication 14 et/ou un procédé comprenant les étapes a) et b), ou a), b) et c) du procédé
- 25 selon la revendication 15.
17. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 16, caractérisée en ce que ladite mycobactérie est une mycobactérie pathogène.
- 30
18. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 17, caractérisée en ce que ladite mycobactérie est une mycobactérie appartenant au groupe du complexe *Mycobacterium tuberculosis*.

19. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 18, caractérisée en ce que ladite mycobactérie est *Mycobacterium tuberculosis*.
- 5 20. Séquence nucléotidique de mycobactérie ou comprenant une séquence nucléotidique de mycobactérie susceptible d'être sélectionnée par un procédé selon l'une des revendications 14 et 15.
- 10 21. Séquence nucléotidique de mycobactérie ou comprenant une séquence nucléotidique de mycobactérie selon la revendication 20, caractérisée en ce que ladite mycobactérie est choisie parmi *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. avium*, *M. leprae*, *M. paratuberculosis*, *M.*
- 15 *kansassi* ou *M. xénopi*.
22. Séquence nucléotidique selon l'une des revendications 20-21 caractérisée en ce qu'elle est choisie parmi les séquences de fragments d'ADN de mycobactérie de séquence
- 20 nucléique SEQ ID N°1 à SEQ ID N° 24C, SEQ ID N°27A à SEQ ID N°27C, SEQ ID N°29 et SEQ ID N° 31A à SEQ ID N°50F.
23. Séquence nucléotidique de mycobactérie l'une des revendications 20-21 caractérisée en ce qu'elle est choisie
- 25 parmi les séquences de fragments d'ADN de mycobactérie de séquence SEQ ID N°1, SEQ ID N°3A, SEQ ID N°5A, SEQ ID N°6A, SEQ ID N°7A, SEQ ID N°8A, SEQ ID N°9A, SEQ ID N°10A, SEQ ID N°27A ou SEQ ID N°29 contenus respectivement dans les vecteurs pDP428 (CNCM, N°I-1818), p6D7 (CNCM, N°I-1814),
- 30 p5F6 (CNCM, N°I-1816), p2A29 (CNCM, N°I-1817), p5B5 (CNCM, N°I-1819), p1C7 (CNCM, N°I-1820), p2D7 (CNCM, N°I-1821), p1B7 (CNCM, N°I-1843), p5A3 (CNCM, N°I-1815) et pM1C25 (CNCM, N°I-2062).
- 35 24. Séquence nucléotidique comprenant la totalité du cadre ouvert de lecture d'une séquence selon l'une quelconque des revendications 20 à 23.

25. Polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi :

- 5 a) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide selon l'une des revendications 20 à 24,
- b) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide selon l'une des revendications 20 à 24,
- 10 c) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide selon l'une des revendications 20 à 24,
- d) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini selon l'une des revendications 20 à 15 24 ou défini en a).

26. Polypeptide, leurs fragments ou fragments biologiquement actifs ou leurs polypeptides homologues, susceptible d'être codé par une séquence nucléotidique de 20 mycobactérie selon l'une des revendications 20 à 25, et susceptible d'être exporté et/ou sécrété, et/ou induit ou réprimé, ou exprimé de façon constitutive lors de l'infection.

- 25 27. Mycobactérie recombinante caractérisée en ce qu'elle est transformée par un vecteur recombinant selon l'une des revendications 1 à 13.

30 28. Polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques de séquence SEQ ID N°1 à SEQ ID N°2.

29. Polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi :

- 35 a) un polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°2,

- b) un polynucléotide dont la séquence nucléique est la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence SEQ ID N°1,
- 5 c) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide défini en a) ou b),
- d) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide défini en a), b) ou c),
- e) un polynucléotide hybridant dans des conditions de forte
10 stringence avec une séquence de polynucléotide défini en a), b), c) ou d),
- f) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini en a), b), c), d) ou e).
- 15 30. Polynucléotide selon l'une des revendications 20 à 25, 28 et 29, caractérisé en ce que sa séquence nucléique hybride avec l'ADN de séquence de mycobactéries et préférentiellement avec de l'ADN de séquences de mycobactéries appartenant au complexe de *Mycobacterium*
20 *tuberculosis*.
31. Polypeptide caractérisé en ce qu'il est codé par une séquence polynucléotidique selon l'une des revendications 20 à 25.
- 25 32. Polypeptide caractérisé en ce qu'il comprend un polypeptide choisi parmi :
- a) un polypeptide dont la séquence d'acides aminés est comprise dans une séquence d'acides aminés choisie parmi
30 les séquences d'acides aminés SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 28 et SEQ ID N° 30 à SEQ ID N° 50F,
- b) un polypeptide homologue au polypeptide défini en a),
- c) un fragment d'au moins 5 acides aminés d'un polypeptide
35 défini en a) ou b),
- d) un fragment biologiquement actif d'un polypeptide défini en a), b), ou c).

33. Polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés SEQ ID N°1 ou SEQ ID N°2, ou polypeptide de séquence d'acides aminés SEQ ID N°28.
34. Polynucléotide caractérisé en ce qu'il code pour un polypeptide selon l'un des revendications 32 et 33.
35. Séquence d'acide nucléique utilisable comme amorce, caractérisée en ce que ladite séquence est choisie parmi les séquences d'acide nucléique de polynucléotide selon l'une des revendications 20 à 25, 28 à 30, et 34.
36. Séquence d'acide nucléique selon la revendication 35, caractérisée en ce que ladite séquence est choisie parmi les séquences SEQ ID N°25 et SEQ ID N°26.
37. Séquence d'acide nucléique selon l'une des revendications 35 et 36 pour la détection et/ou l'amplification de séquences nucléiques.
38. Séquence d'acide nucléique utilisable comme sonde, caractérisée en ce que ladite séquence est choisie parmi les séquences d'acide nucléique selon l'une des revendications 20 à 25, 28 à 30, et 34.
39. Séquence d'acide nucléique selon la revendication 38, caractérisée en ce qu'elle est marquée par un composé radioactif ou par un composé non radioactif.
40. Séquence d'acide nucléique selon l'une des revendications 38 et 39, caractérisée en ce que celle-ci est immobilisée sur un support, de manière covalente ou non-covalente.

41. Séquence d'acide nucléique selon l'une des revendications 38 à 40 pour la détection et/ou l'amplification de séquences nucléiques.
- 5 42. Séquence d'acide nucléique selon l'une des revendications 38 à 41, caractérisée en ce que ladite séquence est une séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence SEQ ID N°1.
- 10 43. Vecteur recombinant de clonage, d'expression et/ou d'insertion, caractérisé en ce qu'il contient une séquence d'acide nucléique de polynucléotide selon l'une des revendications 20 à 25, 28 à 30, et 34.
- 15 44. Cellule hôte, caractérisée en ce qu'elle est transformée par un vecteur recombinant selon la revendication 43.
- 20 45. Cellule hôte selon la revendication 44, caractérisée en ce qu'il s'agit de la souche de *E. coli* transformée par le plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818 ou transformée par le plasmide pM1C25 déposé le 4 août 1998 à la CNCM sous le n°I-2062, ou d'une souche de
- 25 *M. tuberculosis*, *M. bovis* ou *M. africanum* possédant potentiellement tous les systèmes de régulation appropriés.
- 30 46. Procédé de préparation d'un polypeptide, caractérisé en ce qu'il met en oeuvre un vecteur selon la revendication 43.
47. Polypeptide recombinant susceptible d'être obtenu par un procédé selon la revendication 46.
- 35 48. Polypeptide hybride, caractérisé en ce qu'il comporte au moins la séquence d'un polypeptide selon l'une des revendications 26, 32, 33 et 47 et une séquence d'un

polypeptide susceptible d'induire une réponse immunitaire chez l'homme ou l'animal.

49. Polypeptide hybride selon la revendication 48, caractérisé en ce que le polypeptide susceptible d'induire une réponse immunitaire contient au moins un déterminant antigénique capable d'induire une réponse humorale et/ou cellulaire.

50. Polynucléotide codant pour un polypeptide hybride selon l'une des revendications 48 et 49.

51. Polypeptide hybride selon l'une des revendications 48 et 49, caractérisée en ce qu'il s'agit d'une protéine recombinante obtenue par l'expression d'un polynucléotide selon la revendication 50.

52. Procédé pour la détection *in vitro* d'anticorps dirigés contre une mycobactérie et préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec un polypeptide selon l'une des revendications 26, 32, 33 et 47 ;
- b) mise en évidence du complexe antigène-anticorps formé.

53. Procédé pour la détection d'une infection par une mycobactérie et préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère, caractérisé en ce qu'il comprend les étapes suivantes :

- a) préparation d'un échantillon biologique contenant des cellules dudit mammifère plus particulièrement des cellules du système immunitaire dudit mammifère et plus particulièrement encore des cellules T ;

b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'une quelconque des revendications 26, 32, 33 et 47 ;

5 c) détection d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide notamment la prolifération cellulaire et/ou la synthèse de protéines telles que l'interféron gamma;

d) détection d'une réaction d'hypersensibilité retardée ou de sensibilisation du mammifère audit polypeptide.

10

54. Kit pour le diagnostic in vitro d'une infection par une mycobactérie appartenant au complexe *Mycobacterium tuberculosis*, comprenant :

15 a) un polypeptide selon l'une des revendications 26, 32, 33 et 47 ;

b) le cas échéant, les réactifs pour la constitution du milieu propice à la réaction immunologique;

c) les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique ;

20 d) le cas échéant, un échantillon biologique de référence (témoin négatif) dépourvu d'anticorps reconnus par ledit polypeptide ;

25 e) le cas échéant, un échantillon biologique de référence (témoin positif) contenant une quantité prédéterminée d'anticorps reconnus par ledit polypeptide.

30 55. Anticorps mono- ou polyclonaux, leurs fragments, ou anticorps chimériques, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une des revendications 26, 32, 33 et 47 .

56. Anticorps selon la revendication 55, caractérisé en ce qu'il s'agit d'un anticorps marqué.

35 57. Procédé pour la détection spécifique de la présence d'un antigène d'une bactérie du complexe *Mycobacterium*

tuberculosis dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec un anticorps selon l'une des revendications 55 et 56 ;
- 5 b) mise en évidence du complexe antigène-anticorps formé.

58. Kit pour la détection spécifique de la présence d'un antigène d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en

10 ce qu'il comprend les éléments suivants :

- a) un anticorps polyclonal ou monoclonal selon l'une des revendications 55 et 56 ;
- b) les réactifs pour la constitution du milieu propice à la réaction immunologique ;
- 15 c) les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique.

59. Procédé de détection et d'identification rapide d'une mycobactérie et préférentiellement de *M. tuberculosis* dans

20 un échantillon biologique, caractérisé en ce qu'il comporte les étapes suivantes :

- a) isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARN de l'échantillon biologique ;
- 25 b) amplification spécifique de l'ADN des mycobactéries appartenant au complexe *Mycobacterium tuberculosis* à l'aide d'amorces selon l'une des revendications 35 à 37 ;
- c) analyse des produits d'amplification.

30 60. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact d'une sonde oligonucléotidique selon
- 35 l'une des revendications 38 à 42 avec un échantillon biologique, l'ADN contenu dans l'échantillon biologique

ayant, le cas échéant, préalablement été rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;

- 5 b) détection de l'hybride formé entre la sonde oligonucléotidique et l'ADN de l'échantillon biologique.

61. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon
10 biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact d'une sonde oligonucléotidique immobilisée sur un support selon la revendication 40 avec un échantillon biologique, l'ADN de l'échantillon, ayant,
15 le cas échéant, été préalablement rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;

- b) mise en contact de l'hybride formé entre la sonde
20 oligonucléotidique immobilisée sur un support et l'ADN contenu dans l'échantillon biologique, le cas échéant après élimination de l'ADN de l'échantillon biologique n'ayant pas hybridé avec la sonde, avec une sonde oligonucléotidique marquée selon la revendication 39.

25

62. Procédé de détection selon la revendication 61, caractérisé en ce que, préalablement à l'étape a), l'ADN de l'échantillon biologique, ou l'ADNc obtenu par transcription inverse de l'ARN de l'échantillon, est
30 amplifié à l'aide d'un couple d'amorces selon l'une des revendications 35 à 37.

63. Procédé pour la détection de la présence d'une bactérie appartenant au complexe de *Mycobacterium tuberculosis* dans
35 un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec un couple d'amorces selon l'une des revendications 35 à 37, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;
- b) amplification de l'ADN de la bactérie du complexe *Mycobacterium tuberculosis* ;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique marquée selon la revendication 39.

64. Procédé pour la détection de la présence d'une bactérie appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec deux couples d'amorces selon l'une des revendications 35 à 37, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;
- b) amplification de l'ADN de la bactérie du complexe *Mycobacterium tuberculosis* ;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par lesdites amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique marquée selon la revendication 39.

65. Kit pour la détection de la présence d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

a) une sonde oligonucléotidique selon l'une des revendications 38 à 42 ;

b) les réactifs nécessaires à la mise en oeuvre d'une réaction d'hybridation ;

5 c) le cas échéant, un couple d'amorces selon l'une des revendications 35 à 37 ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN (ADN génomique, plasmidique ou ADNc) d'une bactérie du complexe *Mycobacterium tuberculosis*.

10

66. Kit ou nécessaire pour la détection de la présence d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

15 a) une sonde oligonucléotidique, dite sonde de capture, selon la revendication 40 ;

b) une sonde oligonucléotidique, dite sonde de révélation, selon l'une des revendications 38 à 42 ;

20 c) le cas échéant, un couple d'amorces selon l'une des revendications 35 à 37 ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis*.

25 67. Kit pour l'amplification de l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* présent dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

a) un couple d'amorces selon l'une des revendications 35 à 37 ;

30 b) les réactifs nécessaires pour effectuer une réaction d'amplification d'ADN ;

c) éventuellement un composant permettant de vérifier la séquence du fragment amplifié, plus particulièrement une sonde oligonucléotidique selon l'une des revendications 38
35 à 42.

68. Composition immunogène caractérisée en ce qu'elle comprend un ou plusieurs polypeptides selon l'une des revendications 26, 32, 33 et 47 et/ou un ou plusieurs polypeptides hybrides selon l'une des revendications 48, 49 et 51.

69. Vaccin caractérisé en ce qu'il contient un ou plusieurs polypeptides selon l'une des revendications 26, 32, 33 et 47 et/ou un ou plusieurs polypeptides hybrides selon l'une des revendications 48, 49 et 51, en association avec un véhicule pharmaceutiquement compatible et, le cas échéant un ou plusieurs adjuvants de l'immunité appropriés.

70. Vaccin destiné à l'immunisation à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, comprenant un vecteur selon la revendication 43 ou un polynucléotide selon la revendication 50, en association avec un véhicule pharmaceutiquement acceptable.

71. Vaccin caractérisé en ce qu'il contient une ou plusieurs séquences nucléotidiques selon l'une des revendications 20 à 24 et/ou un ou plusieurs polynucléotides selon la revendication 25 en association avec un véhicule pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité appropriés.

72. Méthode de criblage de molécules capables d'inhiber la croissance de mycobactéries ou le maintien de mycobactéries dans un hôte, caractérisée en ce que lesdites molécules bloquent la synthèse ou la fonction des polypeptides codés par une séquence nucléotidique selon l'une quelconque des revendications 20 à 24 ou par un polynucléotide selon la revendication 25.

73. Méthode de criblage selon la revendication 72, caractérisée en ce que les molécules sont des anti-messagers ou induisent la synthèse d'anti-messagers.

5 74. Molécules capables d'inhiber la croissance de mycobactéries ou le maintien de mycobactéries dans un hôte, caractérisées en ce que lesdites molécules sont synthétisées d'après la structure des polypeptides codés
10 par une séquence nucléotidique selon l'une quelconque des revendications 20 à 24 ou par un polynucléotide selon la revendication 25.

15

20

1/185

1 GGATCCAGGGAACGTGACC ATG GTC GTA GGG ATG ACT TGA CAGTTTCAACGGGGTGGACCACCGTTGCGC 72
 1 M V V G M T * 7

73 TCAGAAGGCATACGTTGGTGGAAACACGTCGGAAAGCTGGGAGGTGAATCTG ATG GCT GGC GAC CAA GAG CTG 144
 1 M A G D Q E L 7

145 GAA CTG CGG TTC GAC GTT CCT CTT TAC ACG CTT GCC GAG GCA TCG CGG TAC CTG GTG GTT 204
 8 E L R F D V P L Y T L A E A S R Y L V V 27

205 CCC CGC GCC ACC CTG GCT ACG TGG GCT GAC GGC TAC GAG CGT CGG CCG GCC AAC GCA CCG 264
 28 P R A T L A T W A D G Y E R R P A N A P 47

265 GCG GTC CAG GGG CAA CCG ATC GCC TTT GAC GCC TAT TCG GTC GCG CAG CTT TTT GGC GAC 324
 48 A V Q G Q P I A F D A Y S V A Q L F G D 67

325 GTC ACT GGT GCC CGC GTT GCG GGC GTC CAG CCG CAG CGA CAC CAC ATA CGG CCG GTC CGG 384
 68 V T G A R V A G V Q P Q R H H I R P V R 87

385 TTG CGG GGG CCG TTG GGT GGG GTT GGG TGC CTC CGT CAC CCC AGG CAG TTC GCT GGC TAT 444
 88 L R G P L G G V G C L R H P R Q F A G Y 107

445 TTG TCG CAG TAG CGCGACGGCATTTGTCG ATG TCT TGG TAG CTAGCATCCGGTCGGGGGGCCGCTACCAGCG 515
 108 L S Q * M S W * 4

516 CCAGCGCCGGGGCTCCCGGTCCGGGTAGTGC CGCTCGAGTTGGTCGTGGACCAGCA ATG ACT GCG ACC CGG 587
 1 M T A T R 5

588 CGA CTT CGA AAC CGC CAC CGG TTA GAT TCC CCG ACT GCG TCA TCG CCA GGT AAA CCG CCG 647
 6 R L R N R H R L D S P T A S S P G K P P 25

648 GCA CTA ACG CCA GCA ACC AAC CCG TGA AGACCAACCAACGGCACCTGCGCAGGTTGCGGCTCAACCGCATC 718
 26 A L T P A T N P * 34

719 ATG AAC TGC TGG ATT TCG GAC TCC CCG TAC TCT CGC GCA GTG CGT GCC CGC GAG CCT ACC 778
 1 M N C W I S D S P Y S R A V R A R E P T 20

779 GAA GAT CGC GTG CAT GCG TTC GGC GTG GAC CGC ACA GCA CCT GGA GTT GGC GGC GCC GAG 838
 21 E D R V H A F G V D R T A P G V G G A E 40

839 GGC CGA GAT GGC AGG ATG ACG GAT CGT CGG GGG CGG GAA CTC CCA GGC CGC CGG ACC GTC 898
 41 G R D G R M T D R R G R E L P G R R T V 60

899 GCA AAC CCG TCG CAA ACC CGT CGC AAA CCG TAA GGAGTCATCC ATG AAG ACA GGC ACC GCG 959
 61 A N P S Q T R R K P * M K T G T A 6

960 ACG ACG CGG CGC AGG CTG TTG GCA GTA CTG ATC GCC CTC GCG TTG CCG GGG GCC GCC GTT 1019
 7 T T R R R L L A V L I A L A L P G A A V 26

1020 GCG CTG CTG GCC GAA CCA TCA GCG ACC GGC GCG TCG GAC CCG TGC GCG GCC AGC GAA GTG 1079
 27 A L L A E P S A T G A S D P C A A S E V 46

1080 GCG AGG ACG GTC GGT TCG GTC GCC AAG TCG ATG GGC GAC TAC CTG GAT TCA CAC CCA GAG 1139
 47 A R T V G S V A K S M G D Y L D S H P E 66

1140 ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GGG CCG GGG TCG GTC GCA TCG 1199
 67 T N Q V M T A V L Q Q Q V G P G S V A S 86

1200 CTG AAG GCC CAT TTC GAG GCG AAT CCC AAG GTC GCA TCG GAT CC 1243
 87 L K A H F E A N P K V A S D 100

SEQ ID N° 1

FIGURE 1

FEUILLE DE REMPLACEMENT (REGLE 26)

2/185

Insert du clone contenant DP428 et contenu dans seq1

```

1/1                                     31/11
GAT CGC CTT TGA CGC CTA TTC GGT CGC GCA GCT TTT TGG CGA CGT CAC TGG TGC CCG CGT
asp arg leu OPA arg leu phe gly arg ala ala phe trp arg arg his trp cys pro arg
61/21                                     91/31
TGC GGG CGT CCA GCC GCA GCG ACA CCA CAT ACG GCC GGT CCG GTT GCG GGG GCC GTT GGG
cys gly arg pro ala ala ala thr pro his thr ala gly pro val ala gly ala val gly
121/41                                     151/51
TGG GGT TGG GTG CCT CCG TCA CCC CAG GCA GTT CGC TGG CTA TTT GTC GCA GTA GCG CGA
trp gly trp val pro pro ser pro gln ala val arg trp leu phe val ala val ala arg
181/61                                     211/71
CGG CAT TGT CGA TGT CTT GGT AGC TAG CAT CCG GTC GGG GGG CCG CTA CCA GCG CCA GCG
arg his cys arg cys leu gly ser AMB his pro val gly gly pro leu pro ala pro ala
241/81                                     271/91
CCG GGG CTC CCC GGT CCG GGT AGT GCG CGT CGA GTT GGT CGT GGA CCA GCA ATG ACT GCG
pro gly leu pro gly pro gly ser ala arg arg val gly arg gly pro ala met thr ala
301/101                                     331/111
ACC CGG CGA CTT CGA AAC CGC CAC CGG TTA GAT TCC CCG ACT GCG TCA TCG CCA GGT AAA
thr arg arg leu arg asn arg his arg leu asp ser pro thr ala ser ser pro gly lys
361/121                                     391/131
CCG CCG GCA CTA ACG CCA GCA ACC AAC CCG TGA AGA CCA ACC AAC GGC ACC TGC GCA GGT
pro pro ala leu thr pro ala thr asn pro OPA arg pro thr asn gly thr cys ala gly
421/141                                     451/151
TGC GGC TCA ACC GCA TCA TGA ACT GCT GGA TTT CGG ACT CCC CGT ACT CTC GCG CAG TGC
cys gly ser thr ala ser OPA thr ala gly phe arg thr pro arg thr leu ala gln cys
481/161                                     511/171
GTG CCC GCG AGC CTA CCG AAG ATC GCG TGC ATG CGT TCG GCG TGG ACC GCA CAG CAC CTG
val pro ala ser leu pro lys ile ala cys met arg ser ala trp thr ala gln his leu
541/181                                     571/191
GAG TTG GCG GCG CCG AGG GCC GAG ATG GCA GGA TGA CCG ATC GTC GGG GGC GGG AAC TCC
glu leu ala ala pro arg ala glu met ala gly OPA arg ile val gly gly gly asn ser
601/201                                     631/211
CAG GCC GCC GGA CCG TCG CAA ACC CGT CGC AAA CCC GTC GCA AAC CGT AAG GAG TCA TCC
gln ala ala gly pro ser gln thr arg arg lys pro val ala asn arg lys glu ser ser
661/221                                     691/231
ATG AAG ACA GGC ACC GCG ACG ACG CGG CGC AGG CTG TTG GCA GTA CTG ATC GCC CTC GCG
met lys thr gly thr ala thr thr arg arg arg leu leu ala val leu ile ala leu ala
721/241                                     751/251
TTG CCG GGG GCC GCC GTT GCG CTG CTG GCC GAA CCA TCA GCG ACC GGC GCG TCG GAC CCG
leu pro gly ala ala val ala leu leu ala glu pro ser ala thr gly ala ser asp pro
781/261                                     811/271
TGC GCG GCC AGC GAA GTG GCG AGG ACG GTC GGT TCG GTC GCC AAG TCG ATG GGC GAC TAC
cys ala ala ser glu val ala arg thr val gly ser val ala lys ser met gly asp tyr
841/281                                     871/291
CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GGG
leu asp ser his pro glu thr asn gln val met thr ala val leu gln gln gln val gly
901/301                                     931/311
CCG GGG TCG GTC GCA TCG CTG AAG GCC CAT TTC GAG GCG AAT CCC AAG GTC GCA TCG GAT C
pro gly ser val ala ser leu lys ala his phe glu ala asn pro lys val ala ser asp

```

SEQ ID N° 1A'

FIGURE 1A'

3/185

Insert du clone contenant DP428, autre phase de lecture

2/1 32/11
 ATC GCC TTT GAC GCC TAT TCG GTC GCG CAG CTT TTT GGC GAC GTC ACT GGT GCC CGC GTT
 ile ala phe asp ala tyr ser val ala gln leu phe gly asp val thr gly ala arg val
 62/21 92/31
 GCG GGC GTC CAG CCG CAG CGA CAC CAC ATA CGG CCG GTC CGG TTG CGG GGG CCG TTG GGT
 ala gly val gln pro gln arg his his ile arg pro val arg leu arg gly pro leu gly
 122/41 152/51
 GGG GTT GGG TGC CTC CGT CAC CCC AGG CAG TTC GCT GGC TAT TTG TCG CAG TAG CGC GAC
 gly val gly cys leu arg his pro arg gln phe ala gly tyr leu ser gln AMB arg asp
 182/61 212/71
 GGC ATT GTC GAT GTC TTG GTA GCT AGC ATC CGG TCG GGC GGC CGC TAC CAG CGC CAG CGC
 gly ile val asp val leu val ala ser ile arg ser gly gly arg tyr gln arg gln arg
 242/81 272/91
 CGG GGC TCC CCG GTC CGG GTA GTG CGC GTC GAG TTG GTC GTG GAC CAG CAA TGA CTG CGA
 arg gly ser pro val arg val val arg val glu leu val val asp gln gln OPA leu arg
 302/101 332/111
 CCC GGC GAC TTC GAA ACC GCC ACC GGT TAG ATT CCC CGA CTG CGT CAT CGC CAG GTA AAC
 pro gly asp phe glu thr ala thr gly AMB ile pro arg leu arg his arg gln val asn
 362/121 392/131
 CGC CGG CAC TAA CGC CAG CAA CCA ACC CGT GAA GAC CAA CCA ACG GCA CCT GCG CAG GTT
 arg arg his OCH arg gln gln pro thr arg glu asp gln pro thr ala pro ala gln val
 422/141 452/151
 GCG GCT CAA CCG CAT CAT GAA CTG CTG GAT TTC GGA CTC CCC GTA CTC TCG CGC AGT GCG
 ala ala gln pro his his glu leu leu asp phe gly leu pro val leu ser arg ser ala
 482/161 512/171
 TGC CCG CGA GCC TAC CGA AGA TCG CGT GCA TGC GTT CGG CGT GGA CCG CAC AGC ACC TGG
 cys pro arg ala tyr arg arg ser arg ala cys val arg arg gly pro his ser thr trp
 542/181 572/191
 AGT TGG CGG CGC CGA GGG CCG AGA TGG CAG GAT GAC GGA TCG TCG GGG GCG GGA ACT CCC
 ser trp arg arg arg gly pro arg trp gln asp asp gly ser ser gly ala gly thr pro
 602/201 632/211
 AGG CCG CCG GAC CGT CGC AAA CCC GTC GCA AAC CCG TCG CAA ACC GTA AGG AGT CAT CCA
 arg pro pro asp arg arg lys pro val ala asn pro ser gln thr val arg ser his pro
 662/221 692/231
 TGA AGA CAG GCA CCG CGA CGA CGC GGC GCA GGC TGT TGG CAG TAC TGA TCG CCC TCG CGT
 OPA arg gln ala pro arg arg arg gly ala gly cys trp gln tyr OPA ser pro ser arg
 722/241 752/251
 TGC CGG GGG CCG CCG TTG CGC TGC TGG CCG AAC CAT CAG CGA CCG GCG CGT CGG ACC CGT
 cys arg gly pro pro leu arg cys trp pro asn his gln arg pro ala arg arg thr arg
 782/261 812/271
 GCG CGG CCA GCG AAG TGG CGA GGA CGG TCG GTT CGG TCG CCA AGT CGA TGG GCG ACT ACC
 ala arg pro ala lys trp arg gly arg ser val arg ser pro ser arg trp ala thr thr
 842/281 872/291
 TGG ATT CAC ACC CAG AGA CCA ACC AGG TGA TGA CCG CGG TCT TGC AGC AGC AGG TAG GGC
 trp ile his thr gln arg pro thr arg OPA OPA pro arg ser cys ser ser arg AMB gly
 902/301 932/311
 CGG GGT CGG TCG CAT CGC TGA AGG CCC ATT TCG AGG CGA ATC CCA AGG TCG CAT CGG ATC
 arg gly arg ser his arg OPA arg pro ile ser arg arg ile pro arg ser his arg ile

SEQ ID N° 1B'

FIGURE 1B'

FEUILLE DE REMPLACEMENT (REGLE 26)

4/185

Seq1C: Insert du clone DP428, autre phase de lecture

```

3/1                               33/11
TCG CCT TTG ACG CCT ATT CGG TCG CGC AGC TTT TTG GCG ACG TCA CTG GTG CCC GCG TTG
ser pro leu thr pro ile arg ser arg ser phe leu ala thr ser leu val pro ala leu
63/21                               93/31
CGG GCG TCC AGC CGC AGC GAC ACC ACA TAC GGC CGG TCC GGT TGC GGG GGC CGT TGG GTG
arg ala ser ser arg ser asp thr thr tyr gly arg ser gly cys gly gly arg trp val
123/41                               153/51
GGG TTG GGT GCC TCC GTC ACC CCA GGC AGT TCG CTG GCT ATT TGT CGC AGT AGC GCG ACG
gly leu gly ala ser val thr pro gly ser ser leu ala ile cys arg ser ser ala thr
183/61                               213/71
GCA TTG TCG ATG TCT TGG TAG CTA GCA TCC GGT CGG GGG GCC GCT ACC AGC GCC AGC GCC
ala leu ser met ser trp AMB leu ala ser gly arg gly ala ala thr ser ala ser ala
243/81                               273/91
GGG GCT CCC CGG TCC GGG TAG TGC GCG TCG AGT TGG TCG TGG ACC AGC AAT GAC TGC GAC
gly ala pro arg ser gly AMB cys ala ser ser trp ser trp thr ser asn asp cys asp
303/101                               333/111
CCG GCG ACT TCG AAA CCG CCA CCG GTT AGA TTC CCC GAC TGC GTC ATC GCC AGG TAA ACC
pro ala thr ser lys pro pro pro val arg phe pro asp cys val ile ala arg OCH thr
363/121                               393/131
GCC GGC ACT AAC GCC AGC AAC CAA CCC GTG AAG ACC AAC CAA CGG CAC CTG CGC AGG TTG
ala gly thr asn ala ser asn gln pro val lys thr asn gln arg his leu arg arg leu
423/141                               453/151
CGG CTC AAC CGC ATC ATG AAC TGC TGG ATT TCG GAC TCC CCG TAC TCT CGC GCA GTG CGT
arg leu asn arg ile met asn cys trp ile ser asp ser pro tyr ser arg ala val arg
483/161                               513/171
GCC CGC GAG CCT ACC GAA GAT CGC GTG CAT GCG TTC GGC GTG GAC CGC ACA GCA CCT GGA
ala arg glu pro thr glu asp arg val his ala phe gly val asp arg thr ala pro gly
543/181                               573/191
GTT GGC GGC GCC GAG GGC CGA GAT GGC AGG ATG ACG GAT CGT CGG GGG CGG GAA CTC CCA
val gly gly ala glu gly arg asp gly arg met thr asp arg arg gly arg glu leu pro
603/201                               633/211
GGC CGC CGG ACC GTC GCA AAC CCG TCG CAA ACC CGT CGC AAA CCG TAA GGA GTC ATC CAT
gly arg arg thr val ala asn pro ser gln thr arg arg lys pro OCH gly val ile his
663/221                               693/231
GAA GAC AGG CAC CGC GAC GAC GCG GCG CAG GCT GTT GGC AGT ACT GAT CGC CCT CGC GTT
glu asp arg his arg asp asp ala ala gln ala val gly ser thr asp arg pro arg val
723/241                               753/251
GCC GGG GGC CGC CGT TGC GCT GCT GGC CGA ACC ATC AGC GAC CGG CGC GTC GGA CCC GTG
ala gly gly arg arg cys ala ala gly arg thr ile ser asp arg arg val gly pro val
783/261                               813/271
CGC GGC CAG CGA AGT GGC GAG GAC GGT CGG TTC GGT CGC CAA GTC GAT GGG CGA CTA CCT
arg gly gln arg ser gly glu asp gly arg phe gly arg gln val asp gly arg leu pro
843/281                               873/291
GGA TTC ACA CCC AGA GAC CAA CCA GGT GAT GAC CGC GGT CTT GCA GCA GCA GGT AGG GCC
gly phe thr pro arg asp gln pro gly asp asp arg gly leu ala ala ala gly arg ala
903/301                               933/311
GGG GTC GGT CGC ATC GCT GAA GGC CCA TTT CGA GGC GAA TCC CAA GGT CGC ATC GGA TC
gly val gly arg ile ala glu gly pro phe arg gly glu ser gln gly arg ile gly

```

SEQ ID N° 1C'

FIGURE 1C'

FEUILLE DE REMPLACEMENT (REGLE 26)

5/185

Séquence codante DP428 identique à la séquence Rv0203 prédite par Cole et al.
(Nature 393:537-544)

```

1/1                               31/11
ATG AAG ACA GGC ACC GCG ACG ACG CGG CGC AGG CTG TTG GCA GTA CTG ATC GCC CTC GCG
Met lys thr gly thr ala thr thr arg arg arg leu leu ala val leu ile ala leu ala
61/21                               91/31
TTG CCG GGG GCC GCC GTT GCG CTG CTG GCC GAA CCA TCA GCG ACC GGC GCG TCG GAC CCG
leu pro gly ala ala val ala leu leu ala glu pro ser ala thr gly ala ser asp pro
121/41                               151/51
TGC GCG GCC AGC GAA GTG GCG AGG ACG GTC GGT TCG GTC GCC AAG TCG ATG GGC GAC TAC
cys ala ala ser glu val ala arg thr val gly ser val ala lys ser met gly asp tyr
181/61                               211/71
CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GGG
leu asp ser his pro glu thr asn gln val met thr ala val leu gln gln gln val gly
241/81                               271/91
CCG GGG TCG GTC GCA TCG CTG AAG GCC CAT TTC GAG GCG AAT CCC AAG GTC GCA TCG GAT
pro gly ser val ala ser leu lys ala his phe glu ala asn pro lys val ala ser asp
301/101                               331/111
CTG CAC GCG CTT TCG CAA CCG CTG ACC GAT CTT TCG ACT CGG TGC TCG CTG CCG ATC AGC
leu his ala leu ser gln pro leu thr asp leu ser thr arg cys ser leu pro ile ser
361/121                               391/131
GGC CTG CAG GCG ATC GGT TTG ATG CAG GCG GTG CAG GGC GCC CGC CGG TAG
gly leu gln ala ile gly leu met gln ala val gln gly ala arg arg AMB

```

SEQ ID N° 1D

FIGURE 1D

ORF contenant la séquence DP428 et faisant partie de seq1A'

```

1/1                               31/11
TGA CGG ATC GTC GGG GGC GGG AAC TCC CAG GCC GCC GGA CCG TCG CAA ACC CGT CGC AAA
OPA arg ile val gly gly gly asn ser gln ala ala gly pro ser gln thr arg arg lys
61/21                               91/31
CCC GTC GCA AAC CGT AAG GAG TCA TCC ATG AAG ACA GGC ACC GCG ACG ACG CGG CGC AGG
pro val ala asn arg lys glu ser ser met lys thr gly thr ala thr thr arg arg arg
121/41                               151/51
CTG TTG GCA GTA CTG ATC GCC CTC GCG TTG CCG GGG GCC GCC GTT GCG CTG CTG GCC GAA
leu leu ala val leu ile ala leu ala leu pro gly ala ala val ala leu leu ala glu
181/61                               211/71
CCA TCA GCG ACC GGC GCG TCG GAC CCG TGC GCG GCC AGC GAA GTG GCG AGG ACG GTC GGT
pro ser ala thr gly ala ser asp pro cys ala ala ser glu val ala arg thr val gly
241/81                               271/91
TCG GTC GCC AAG TCG ATG GGC GAC TAC CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATG
ser val ala lys ser met gly asp tyr leu asp ser his pro glu thr asn gln val met
301/101                               331/111
ACC GCG GTC TTG CAG CAG CAG GTA GGG CCG GGG TCG GTC GCA TCG CTG AAG GCC CAT TTC
thr ala val leu gln gln gln val gly pro gly ser val ala ser leu lys ala his phe
361/121                               391/131
GAG GCG AAT CCC AAG GTC GCA TCG GAT CTG CAC GCG CTT TCG CAA CCG CTG ACC GAT CTT
glu ala asn pro lys val ala ser asp leu his ala leu ser gln pro leu thr asp leu
421/141                               451/151
TCG ACT CGG TGC TCG CTG CCG ATC AGC GGC CTG CAG GCG ATC GGT TTG ATG CAG GCG GTG
ser thr arg cys ser leu pro ile ser gly leu gln ala ile gly leu met gln ala val
481/161
CAG GGC GCC CGC CGG TAG
gln gly ala arg arg AMB

```

SEQ ID N° 1F

6/185

491 CCGGTCGGGGGGCCGCTACCAGCGCCAGCGCGGGGCTCCCGGTCGGGTA GTG CGC GTC GAG TTG GTC GTG 563
 1 V R V E L V V 7

564 GAC CAG CAA TGA CTGCGACCCGGGACTTCGAAACCGCCACCGGTTAGATTCCCCGACTGCGTCATCGCCAGGTAA 639
 8 D Q Q *

640 ACCGCGGCACTAACGCCAGCAACCAACCC GTG AAG ACC AAC CAA CGG CAC CTG CGC AGG TTG CGG 705
 1 V K T N Q R H L R R L R 12

706 CTC AAC CGC ATC ATG AAC TGC TGG ATT TCG GAC TCC CCG TAC TCT CGC GCA GTG CGT GCC 765
 13 L N R I M N C W I S D S P Y S R A V R A 32

766 CGC GAG CCT ACC GAA GAT CGC GTG CAT GCG TTC GGC GTG GAC CGC ACA GCA CCT GGA GTT 825
 33 R E P T E D R V H A F G V D R T A P G V 52

826 GGC GGC GCC GAG GGC CGA GAT GGC AGG ATG ACG GAT CGT CGG GGG CGG GAA CTC CCA GGC 885
 53 G G A E G R D G R M T D R R G R E L P G 72

886 CGC CGG ACC GTC GCA AAC CCG TCG CAA ACC CGT CGC AAA CGG TAA GGAGTCATCC ATG AAG 946
 73 R R T V A N P S Q T R R K P * xxxxxx M K 2

947 ACA GGC ACC GCG ACG ACG CGG CGC AGG CTG TTG GCA GTA CTG ATC GCC CTC GCG TTG CCG 1006
 3 T G T A T T R R R L L A V L I A L A L P 22

1007 GGC GCC GCC GTT GCG CTG CTG GCC GAA CCA TCA GCG ACC GGC GCG TCG GAC CCG TGC GCG 1066
 23 G A A V A L L A E P S A T G A S D P C A 42

1067 GCC AGC GAA GTG GCG AGG ACG GTC GGT TCG GTC GCC AAG TCG ATG GGC GAC TAC CTG GAT 1126
 43 A S E V A R T V G S V A K S M G D Y L D 62

1127 TCA CAC CCA GAG ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GGG CCG GGG 1186
 63 S H P E T N Q V M T A V L Q Q Q V G P G 82

1187 TCG GTC GCA TCG CTG AAG GCC CAT TTC GAG GCG AAT CCC AAG GTC GCA TCG GAT CTG CAC 1246
 83 S V A S L K A H F E A N P K V A S D L H 102

1247 GCG CTT TCG CAA CCG CTG ACC GAT CTT TCG ACT CCG TGC TCG CTG CCG ATC AGC GGC CTG 1306
 103 A L S Q P L T D L S T R C S L P I S G L 122

1307 CAG GCG ATC GGT TTG ATG CAG GCG GTG CAG GGC GCC CGC CGG TAG ATG CCG GAC CGC CGC 1366
 123 Q A I G L M Q A V Q G A R R * M P D R R 5

1367 CGG GTC CGG CGC AGT CGA CGT GAG GCA GCG GTC GCC TAC CGG GGC GGT GTC TCG CCG CCT 1426
 6 R V R R S R R E A A V A Y R G G V S P P 25

1427 TCT GGT CGC AGG TCA GGG GTC GGC GCT GGA CCT TGC GGT GTG GTT TCG ACC GGG TCG TCG 1486
 26 S G R R S G V G A G P C G V V S T G S S 45

1487 CAG GGT GTG CCC TGC GGT TGG ATG ACA AGT CGC AGG TTT GGA TCG GTT GGC GGG TCG CGA 1546
 46 Q G V P C G W M T S R R F G S V G G S R 65

1547 TCG TTG T 1553
 66 S L 67

SEQ ID N° 2

FIGURE 2

FEUILLE DE REMPLACEMENT (REGLE 26)

7/185

31/11
 TCG CCG GCT CGC GGA CGT AGA TAA TAG CTC ACC GTT GGA CGA CCT CGA CAG GGT CCT TTG
 ser pro ala arg gly arg arg OCH AMB leu thr val gly arg pro arg gln gly pro leu
 61/21
 TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG TCG GGT CAT CGC CTA AGG CTA CCG TTC
 OPA leu pro gly leu thr arg thr thr thr glu ser gly his arg leu arg leu pro phe
 121/41
 TGA CCT GGG GTG CGT GGG CGC CGA CGA GTG AGG CAG TCA TGT CTC AGG GCC CAC CGC CAC
 OPA pro gly val arg gly arg arg arg val arg gln ser cys leu arg ala his arg his
 181/61
 CTC GGT CGC CGG CAG TGT CAG CAT GTG CAG ATG ACT CCA CGC AGC TTG TTC GTG TTG GTG
 leu gly arg arg gln cys gln his val gln met thr pro arg ser leu phe val leu val
 241/81
 TCG TGG TTG CGA CGA CTT GGC GCT GGT GAG CGC ACC CGC CGG CGT CGT GCC GCG CAT GCG
 ser trp leu arg arg leu gly ala gly glu arg thr arg arg arg arg ala ala his ala
 301/101
 GAT C
 asp

SEQ ID N° 3A

FIGURE 3A

32/11
 CGC CGG CTC GCG GAC GTA GAT AAT AGC TCA CCG TTG GAC GAC CTC GAC AGG GTC CTT TGT
 arg arg leu ala asp val asp asn ser ser pro leu asp asp leu asp arg val leu cys
 62/21
 GAC TGC CGG GCT TGA CGC GGA CGA CCA CAG AGT CGG GTC ATC GCC TAA GGC TAC CGT TCT
 asp cys arg ala OPA arg gly arg pro gln ser arg val ile ala OCH gly tyr arg ser
 122/41
 GAC CTG GGG TGC GTG GGC GCC GAC GAG TGA GGC AGT CAT GTC TCA GGG CCC ACC GCC ACC
 asp leu gly cys val gly ala asp glu OPA gly ser his val ser gly pro thr ala thr
 182/61
 TCG GTC GCC GGC AGT GTC AGC ATG TGC AGA TGA CTC CAC GCA GCT TGT TCG TGT TGG TGT
 ser val ala gly ser val ser met cys arg OPA leu his ala ala cys ser cys trp cys
 242/81
 CGT GGT TGC GAC GAC TTG GCG CTG GTG AGC GCA CCC GCC GGC GTC GTG CCG CGC ATG CGG
 arg gly cys asp asp leu ala leu val ser ala pro ala gly val val pro arg met arg
 302/101
 ATC
 ile

SEQ ID N° 3B

FIGURE 3B

8/185

33/11
GCC GGC TCG CGG ACG TAG ATA ATA GCT CAC CGT TGG ACG ACC TCG ACA GGG TCC TTT GTG
ala gly ser arg thr AMB ile ile ala his arg trp thr thr ser thr gly ser phe val
63/21
93/31
ACT GCC GGG CTT GAC GCG GAC GAC CAC AGA GTC GGG TCA TCG CCT AAG GCT ACC GTT CTG
thr ala gly leu asp ala asp asp his arg val gly ser ser pro lys ala thr val leu
123/41
153/51
ACC TGG GGT GCG TGG GCG CCG ACG AGT GAG GCA GTC ATG TCT CAG GGC CCA CCG CCA CCT
thr trp gly ala trp ala pro thr ser glu ala val met ser gln gly pro pro pro pro
183/61
213/71
CGG TCG CCG GCA GTG TCA GCA TGT GCA GAT GAC TCC ACG CAG CTT GTT CGT GTT GGT GTC
arg ser pro ala val ser ala cys ala asp ser thr gln leu val arg val gly val
243/81
273/91
GTG GTT GCG ACG ACT TGG CGC TGG TGA GCG CAC CCG CCG GCG TCG TGC CGC GCA TGC GGA
val val ala thr thr trp arg trp OPA ala his pro pro ala ser cys arg ala cys gly

TC

SEQ ID N° 3C

FIGURE 3C

31/11
CCA ATT TTC CTT CGC GCC GTG CAA TAC CAT CTG CAA GAC CAG CGA CGG CCC GTG GTT GCG
pro ile phe leu arg ala val gln tyr his leu gln asp gln arg arg pro val val ala
61/21
91/31
GTC GCG CAG CTT GCG GAA ACC GGG TAT GGA CCC TGC CGT ACC GTT GTT GCC ACT TGA TGT
val ala gln leu ala glu thr gly tyr gly pro cys arg thr val val ala thr OPA cys
121/41
151/51
CGT CGC TCT CCA CCC GTC GGG GGG CGA AAG CCA TTC CGA CAC TGG GAT CCT CAA AAC GTC
arg arg ser pro pro val gly gly arg lys pro phe arg his trp asp pro gln asn val
181/61
211/71
GGC TGA GTG TCT GCA GGG CTC CGG GGA GCA GCC GAT CAT CAC CAT GTA CGA ACT GAA TAA
gly OPA val ser ala gly leu arg gly ala ala asp his his his val arg thr glu OCH
241/81
271/91
GTC CCC CGC GCG CGA CTT CCA GAC ATT TGT TGT GGT TTC GGT TGA GGC CGA GGC GAG GCT
val pro arg ala arg leu pro asp ile cys cys gly phe gly OPA gly arg gly glu ala
301/101
331/111
CAT TTC GCA GCA ACC GGT CTC CGG GTC GCA GCA TCG TTG CGG CGA TCG CGG CGC AGT CGT
his phe ala ala thr gly leu arg val ala ala ser leu arg arg ser arg arg ser arg
361/121
CGG ACG AGT CGT CGT CAA CGA CCA CGA TC
arg thr ser arg arg gln arg pro arg

SEQ ID N° 4A

FIGURE 4A

9/185

32/11
 CAA TTT TCC TTC GCG CCG TGC AAT ACC ATC TGC AAG ACC AGC GAC GGC CCG TGG TTG CCG
 gln phe ser phe ala pro cys asn thr ile cys lys thr ser asp gly pro trp leu arg
 62/21
 TCG CGC AGC TTG CGG AAA CCG GGT ATG GAC CCT GCC GTA CCG TTG TTG CCA CTT GAT GTC
 ser arg ser leu arg lys pro gly met asp pro ala val pro leu leu pro leu asp val
 122/41
 GTC GCT CTC CAC CCG TCG GGG GGC GAA AGC CAT TCC GAC ACT GGG ATC CTC AAA ACG TCG
 val ala leu his pro ser gly gly glu ser his ser asp thr gly ile leu lys thr ser
 182/61
 GCT GAG TGT CTG CAG GGC TCC GGG GAG CAG CCG ATC ATC ACC ATG TAC GAA CTG AAT AAG
 ala glu cys leu gln gly ser gly glu gln pro ile ile thr met tyr glu leu asn lys
 242/81
 TCC CCC GCG CGC GAC TTC CAG ACA TTT GTT GTG GTT TCG GTT GAG GCC GAG GCG AGG CTC
 ser pro ala arg asp phe gln thr phe val val val ser val glu ala glu ala arg leu
 302/101
 ATT TCG CAG CAA CCG GTC TCC GGG TCG CAG CAT CGT TGC GGC GAT CGC GGC GCA GTC GTC
 ile ser gln gln pro val ser gly ser gln his arg cys gly asp arg gly ala val val
 362/121
 GGA CGA GTC GTC GTC AAC GAC CAC GAT C
 gly arg val val val asn asp his asp

SEQ ID N° 4B

FIGURE 4B

33/11
 AAT TTT CCT TCG CGC CGT GCA ATA CCA TCT GCA AGA CCA GCG ACG GCC CGT GGT TGC GGT
 asn phe pro ser arg arg ala ile pro ser ala arg pro ala thr ala arg gly cys gly
 63/21
 CGC GCA GCT TGC GGA AAC CGG GTA TGG ACC CTG CCG TAC CGT TGT TGC CAC TTG ATG TCG
 arg ala ala cys gly asn arg val trp thr leu pro tyr arg cys cys his leu met ser
 123/41
 TCG CTC TCC ACC CGT CGG GGG GCG AAA GCC ATT CCG ACA CTG GGA TCC TCA AAA CGT CGG
 ser leu ser thr arg arg gly ala lys ala ile pro thr leu gly ser ser lys arg arg
 183/61
 CTG AGT GTC TGC AGG GCT CCG GGG AGC AGC CGA TCA TCA CCA TGT ACG AAC TGA ATA AGT
 leu ser val cys arg ala pro gly ser ser arg ser ser pro cys thr asn OPA ile ser
 243/81
 CCC CCG CGC GCG ACT TCC AGA CAT TTG TTG TGG TTT CGG TTG AGG CCG AGG CGA GGC TCA
 pro pro arg ala thr ser arg his leu leu trp phe arg leu arg pro arg arg gly ser
 303/101
 TTT CGC AGC AAC CGG TCT CCG GGT CGC AGC ATC GTT GCG GCG ATC GCG GCG CAG TCG TCG
 phe arg ser asn arg ser pro gly arg ser ile val ala ala ile ala ala gln ser ser
 363/121
 GAC GAG TCG TCG TCA ACG ACC ACG ATC
 asp glu ser ser ser thr thr thr ile

SEQ ID N° 4C

FIGURE 4C
 FEUILLE DE REMPLACEMENT (REGLE 26)

10/185

partie de la séquence nucléotidique de seq4A

```

1/1                               31/11
CCG CGC GCG ACT TCC AGA CAT TTG TTG TGG TTT CGG TTG AGG CCG AGG CGA GGC TCA TTT
pro arg ala thr ser arg his leu leu trp phe arg leu arg pro arg arg gly ser phe
61/21                               91/31
CGC AGC AAG CGG TCT CCG GGT CGC AGC ATC GTT GCG GCG ATC GCG GCG CAG TCG TCG GAC
arg ser lys arg ser pro gly arg ser ile val ala ala ile ala ala gln ser ser asp
121/41
GAG TCG TCG TCA ACG ACC ACG ATC
glu ser ser ser thr thr thr ile

```

SEQ ID N° 4A'

FIGURE 4A'

```

1/1                               31/11
CGC GCG CGA CTT CCA GAC ATT TGT TGT GGT TTC GGT TGA GGC CGA GGC GAG GCT CAT TTC
arg ala arg leu pro asp ile cys cys gly phe gly OPA gly arg gly glu ala his phe
61/21                               91/31
GCA GCA AGC GGT CTC CGG GTC GCA GCA TCG TTG CGG CGA TCG CGG CGC AGT CGT CGG ACG
ala ala ser gly leu arg val ala ala ser leu arg arg ser arg arg ser arg arg thr
121/41
AGT CGT CGT CAA CGA CCA CGA TC
ser arg arg gln arg pro arg

```

SEQ ID N° 4B'

FIGURE 4B'

```

1/1                               31/11
GCC GCG CGC GAC TTC CAG ACA TTT GTT GTG GTT TCG GTT GAG GCC GAG GCG AGG CTC ATT
ala ala arg asp phe gln thr phe val val val ser val glu ala glu ala arg leu ile
61/21                               91/31
TCG CAG CAA GCG GTC TCC GGG TCG CAG CAT CGT TGC GGC GAT CGC GGC GCA GTC GTC GGA
ser gln gln ala val ser gly ser gln his arg cys gly asp arg gly ala val val gly
121/41
CGA GTC GTC GTC AAC GAC CAC GAT C
arg val val val asn asp his asp

```

SEQ ID N° 4C'

FIGURE 4C'

11/185

ORF d'après par Cole et al. (Nature 393:537-544) et contenant la séquence Seq 4A'

1/1

31/11

tga ata agt ccg ccg cgc gcg act tcc aga cat ttg ttg tgg ttt cgg ttg agg ccg agg
OPA ile ser pro pro arg ala thr ser arg his leu leu trp phe arg leu arg pro arg
61/21

91/31

cga ggc tca ttt cgc agc aag cgg tct ccg ggt cgc agc atc gtt gcg gcg atc gcg gcg
arg gly ser phe arg ser lys arg ser pro gly arg ser ile val ala ala ile ala ala
121/41

151/51

cag tcg tcg gac gag tcg tcg tca acg acc acg atc tcg aac tcg acg ccc tcc tgt tcg
gln ser ser asp glu ser ser ser thr thr thr ile ser asn ser thr pro ser cys ser
181/61

211/71

agg atg cta cgc aga cag cgc tcg atg gtg gcg ccg ttg ttg tac atc ggg atg cac acc
arg met leu arg arg gln arg ser met val ala pro leu leu tyr ile gly met his thr
241/81

271/91

gag ata agc ggt ttc gcc ggg ttc acc gat acc acg ctt gat gca tca cca ggc acc aca
glu ile ser gly phe ala gly phe thr asp thr thr leu asp ala ser pro gly thr thr
301/101

tgg cga ctc aga gac tag

trp arg leu arg asp AMB

SEQ ID N° 4F

FIGURE 4F

séquence en amont de seq4A' et en fusion avec seq4A'

1/1

31/11

GCA ACC TAC CAG CAG AGC CAG GGG CTC ACA GGA CCT AAA GGA GTA GCG CCC ATG GCT GAT
ala thr tyr gln gln ser gln gly leu thr gly pro lys gly val ala pro met ala asp

C

SEQ ID N° 4J

FIGURE 4J

seq4J' dans une autre phase de lecture

1/1

31/11

ACG CAA CCT ACC AGC AGA GCC AGG GGC TCA CAG GAC CTA AAG GAG TAG CGC CCA TGG CTG
thr gln pro thr ser arg ala arg gly ser gln asp leu lys glu AMB arg pro trp leu
61/21

ATC

ile

SEQ ID N° 4K

FIGURE 4K

seq 4J' dans la troisième phase de lecture

1/1

31/11

CGC AAC CTA CCA GCA GAG CCA GGG GCT CAC AGG ACC TAA AGG AGT AGC GCC CAT GGC TGA
arg asn leu pro ala glu pro gly ala his arg thr OCH arg ser ser ala his gly OPA

TC

SEQ ID N° 4L

FIGURE 4L

FEUILLE DE REMPLACEMENT (REGLE 26)

12/185

séquence Rv2050 prédite par Cole et al. (Nature 393:537-544) et contenant seq4J
1/1 31/11
ATG GCT GAT CGT GTC CTG AGG GGC AGT CGC CTC GGA GCC GTG AGC TAT GAG ACC GAC CGC
Met ala asp arg val leu arg gly ser arg leu gly ala val ser tyr glu thr asp arg
61/21 91/31
AAC CAC GAC CTG GCG CCG CGC CAG ATC GCG CGG TAC CGC ACC GAC AAC GGC GAG GAG TTC
asn his asp leu ala pro arg gln ile ala arg tyr arg thr asp asn gly glu glu phe
121/41 151/51
GAA GTC CCG TTC GCC GAT GAC GCC GAG ATC CCC GGC ACC TGG TTG TGC CGC AAC GGC ATG
glu val pro phe ala asp asp ala glu ile pro gly thr trp leu cys arg asn gly met
181/61 211/71
GAA GGC ACC CTG ATC GAG GGC GAC CTG CCC GAG CCG AAG AAG GTT AAG CCG CCC CGG ACG
glu gly thr leu ile glu gly asp leu pro glu pro lys lys val lys pro pro arg thr
241/81 271/91
CAC TGG GAC ATG CTG CTG GAG CGC CGT TCC ATC GAA GAA CTC GAA GAG TTA CTT AAG GAG
his trp asp met leu leu glu arg arg sér ile glu glu leu glu glu leu leu lys glu
301/101 331/111
CGC CTC GAG CTC ATT CGG TCA CGT CGG CGC GGC TGA
arg leu glu leu ile arg ser arg arg arg gly OPA

SEQ ID N° 4M

FIGURE 4M

ORF d'après par Cole et al. (Nature 393:537-544) et contenant la séquence Rv2050
1/1 31/11
TAG TCC GCC CGG GTG TCC GAT CCC GGT ATC ATT GAT GGT CGC GCC GCG CGC GTC GCG TGC
AMB ser ala arg val ser asp pro gly ile ile asp gly arg ala ala arg val ala cys
61/21 91/31
CGG GAA CTA CGC AGA CGG CCG CAG CGT TTG CCA ACC GGA GCC AGT CGC CAG TAC GCA ACC
arg glu leu arg arg arg pro gln arg leu pro thr gly ala ser arg gln tyr ala thr
121/41 151/51
TAC CAG CAG AGC CCA GGG CTC ACA GGA CCT AAA GGA GTA GCG CCC ATG GCT GAT CGT GTC
tyr gln gln ser pro gly leu thr gly pro lys gly val ala pro met ala asp arg val
181/61 211/71
CTG AGG GGC AGT CGC CTC GGA GCC GTG AGC TAT GAG ACC GAC CGC AAC CAC GAC CTG GCG
leu arg gly ser arg leu gly ala val ser tyr glu thr asp arg asn his asp leu ala
241/81 271/91
CCG CGC CAG ATC GCG CGG TAC CGC ACC GAC AAC GGC GAG GAG TTC GAA GTC CCG TTC GCC
pro arg gln ile ala arg tyr arg thr asp asn gly glu glu phe glu val pro phe ala
301/101 331/111
GAT GAC GCC GAG ATC CCC GGC ACC TGG TTG TGC CGC AAC GGC ATG GAA GGC ACC CTG ATC
asp asp ala glu ile pro gly thr trp leu cys arg asn gly met glu gly thr leu ile
361/121 391/131
GAG GGC GAC CTG CCC GAG CCG AAG AAG GTT AAG CCG CCC CGG ACG CAC TGG GAC ATG CTG
glu gly asp leu pro glu pro lys lys val lys pro pro arg thr his trp asp met leu
421/141 451/151
CTG GAG CGC CGT TCC ATC GAA GAA CTC GAA GAG TTA CTT AAG GAG CGC CTC GAG CTC ATT
leu glu arg arg ser ile glu glu leu glu glu leu leu lys glu arg leu glu leu ile
481/161
CGG TCA CGT CGG CGC GGC TGA
arg ser arg arg arg gly OPA

SEQ ID N° 4N

FEUILLE DE REMPLACEMENT (REGLE 26)

13/185

31/11
 GAT CGC GGT CAA CGA GGC CGA ATA CGG CGA GAT GTG GGC CCA AGA CGC CGC CGC GAT GTT
 asp arg gly gln arg gly arg ile arg arg asp val gly pro arg arg arg arg asp val
 61/21
 TGG CTA CGC CGC GGC GAC GGC GAC GGC GAC GGC GAC GTT GCT GCC GTT CGA GGA GGC GCC
 trp leu arg arg gly asp gly asp gly asp gly asp val ala ala val arg gly gly ala
 121/41
 GGA GAT GAC CAG CGC GGG TGG GCT CCT CGA GCA GGC CGC CGC GGT CGA GGA GGC CTC CGA
 gly asp asp gln arg gly trp ala pro arg ala gly arg arg gly arg gly gly leu arg
 181/61
 CAC CGC CGC GGC GAA CCA GTT GAT GAA CAA TGT GCC CCA GGC GCT GCA ACA GCT GGC CCA
 his arg arg gly glu pro val asp glu gln cys ala pro gly ala ala thr ala gly pro
 241/81
 GCC CAC GCA GGG CAC CAC GCC TTC TTC CAA GCT GGG TGG CCT GTG GAA GAC GGT CTC GCC
 ala his ala gly his his ala phe phe gln ala gly trp pro val glu asp gly leu ala
 301/101
 GCA TCG GTC GCC GAT C
 ala ser val ala asp

SEQ ID N° 5A

FIGURE 5A

32/11
 ATC GCG GTC AAC GAG GCC GAA TAC GGC GAG ATG TGG GCC CAA GAC GCC GCC GCG ATG TTT
 ile ala val asn glu ala glu tyr gly glu met trp ala gln asp ala ala ala met phe
 62/21
 GGC TAC GCC GCG GCG ACG GCG ACG GCG ACG GCG ACG TTG CTG CCG TTC GAG GAG GCG CCG
 gly tyr ala ala ala thr ala thr ala thr ala thr leu leu pro phe glu glu ala pro
 122/41
 GAG ATG ACC AGC GCG GGT GGG CTC CTC GAG CAG GCC GCC GCG GTC GAG GAG GCC TCC GAC
 glu met thr ser ala gly gly leu leu glu gln ala ala ala val glu glu ala ser asp
 182/61
 ACC GCC GCG GCG AAC CAG TTG ATG AAC AAT GTG CCC CAG GCG CTG CAA CAG CTG GCC CAG
 thr ala ala ala asn gln leu met asn asn val pro gln ala leu gln gln leu ala gln
 242/81
 CCC ACG CAG GGC ACC ACG CCT TCT TCC AAG CTG GGT GGC CTG TGG AAG ACG GTC TCG CCG
 pro thr gln gly thr thr pro ser ser lys leu gly gly leu trp lys thr val ser pro
 302/101
 CAT CGG TCG CCG ATC
 his arg ser pro ile

SEQ ID N° 5B

FIGURE 5B

14/185

33/11
 TCG CGG TCA ACG AGG CCG AAT ACG GCG AGA TGT GGG CCC AAG ACG CCG CCG CGA TGT TTG
 ser arg ser thr arg pro asn thr ala arg cys gly pro lys thr pro pro arg cys leu
 63/21
 GCT ACG CCG CGG CGA CGG CGA CGG CGA CGG CGA CGT TGC TGC CGT TCG AGG AGG CGC CGG
 ala thr pro arg arg arg arg arg arg arg arg arg arg cys cys arg ser arg arg arg
 123/41
 AGA TGA CCA GCG CGG GTG GGC TCC TCG AGC AGG CCG CCG CGG TCG AGG AGG CCT CCG ACA
 arg OPA pro ala arg val gly ser ser ser arg pro pro arg ser arg arg pro pro thr
 183/61
 CCG CCG CGG CGA ACC AGT TGA TGA ACA ATG TGC CCC AGG CGC TGC AAC AGC TGG CCC AGC
 pro pro arg arg thr ser OPA OPA thr met cys pro arg arg cys asn ser trp pro ser
 243/81
 CCA CGC AGG GCA CCA CGC CTT CTT CCA AGC TGG GTG GCC TGT GGA AGA CGG TCT CGC CGC
 pro arg arg ala pro arg leu leu pro ser trp val ala cys gly arg arg ser arg arg
 303/101
 ATC GGT CGC CGA TC
 ile gly arg arg

SEQ ID N° 5C

FIGURE 5C

partie de la séquence nucléotidique Seq 5A

1/1
 CGC CGC GGC GAC GGC GAC GGC GAC GGC GAC GTT GCT GCC GTT CGA GGA GGC GCC GGA GAT
 arg arg gly asp gly asp gly asp gly asp val ala ala val arg gly gly ala gly asp
 61/21
 GAC CAG CGC GGG TGG GCT CCT CGA GCA GGC CGC CGC GGT CGA GGA GGC CTC CGA CAC CGC
 asp gln arg gly trp ala pro arg ala gly arg arg gly arg gly gly leu arg his arg
 121/41
 CGC GGC GAA CCA GTT GAT GAA CAA TGT GCC CCA GGC GCT GCA ACA GCT GGC CCA GCC CAC
 arg gly glu pro val asp glu gln cys ala pro gly ala ala thr ala gly pro ala his
 181/61
 GCA GGG CAC CAC GCC TTC TTC CAA GCT GGG TGG CCT GTG GAA GAC GGT CTC GCC GCA TCG
 ala gly his his ala phe phe gln ala gly trp pro val glu asp gly leu ala ala ser
 241/81
 GTC GCC GAT C
 val ala asp

SEQ ID N° 5A'

FIGURE 5A'

15/185

```

1/1                               31/11
TAC GCC GCG GCG ACG GCG ACG GCG ACG GCG ACG TTG CTG CCG TTC GAG GAG GCG CCG GAG
tyr ala ala ala thr ala thr ala thr ala thr leu leu pro phe glu glu ala pro glu
61/21                               91/31
ATG ACC AGC GCG GGT GGG CTC CTC GAG CAG GCC GCC GCG GTC GAG GAG GCC TCC GAC ACC
met thr ser ala gly gly leu leu glu gln ala ala ala val glu glu ala ser asp thr
121/41                               151/51
GCC GCG GCG AAC CAG TTG ATG AAC AAT GTG CCC CAG GCG CTG CAA CAG CTG GCC CAG CCC
ala ala ala asn gln leu met asn asn val pro gln ala leu gln gln leu ala gln pro
181/61                               211/71
ACG CAG GGC ACC ACG CCT TCT TCC AAG CTG GGT GGC CTG TGG AAG ACG GTC TCG CCG CAT
thr gln gly thr thr pro ser ser lys leu gly gly leu trp lys thr val ser pro his
241/81
CGG TCG CCG ATC
arg ser pro ile

```

SEQ ID N° 5B'

FIGURE 5B'

```

1/1                               31/11
ACG CCG CGG CGA CGG CGA CGG CGA CGG CGA CGT TGC TGC CGT TCG AGG AGG CGC CGG AGA
thr pro arg arg arg arg arg arg arg arg arg cys cys arg ser arg arg arg arg arg
61/21                               91/31
TGA CCA GCG CGG GTG GGC TCC TCG AGC AGG CCG CCG CGG TCG AGG AGG CCT CCG ACA CCG
OPA pro ala arg val gly ser ser ser arg pro pro arg ser arg arg pro pro thr pro
121/41                               151/51
CCG CGG CGA ACC AGT TGA TGA ACA ATG TGC CCC AGG CGC TGC AAC AGC TGG CCC AGC CCA
pro arg arg thr ser OPA OPA thr met cys pro arg arg cys asn ser trp pro ser pro
181/61                               211/71
CGC AGG GCA CCA CGC CTT CTT CCA AGC TGG GTG GCC TGT GGA AGA CGG TCT CGC CGC ATC
arg arg ala pro arg leu leu pro ser trp val ala cys gly arg arg ser arg arg ile
241/81
GGT CGC CGA TC
gly arg arg

```

SEQ ID N° 5C'

FIGURE 5C'

16/185

ORF prédite par Cole et al. (Nature 393:537-544) et contenant seq5A'

```

1/1                                     31/11
tga act gat gat tct gat agc gac caa cct ctt ggg gca aaa cac ccc ggc gat cgc ggt
OPA thr asp asp ser asp ser asp gln pro leu gly ala lys his pro gly asp arg gly
61/21                                     91/31
caa cga ggc cga ata cgg cga gat gtg ggc cca aga cgc cgc cgc gat gtt tgg cta cgc
gln arg gly arg ile arg arg asp val gly pro arg arg arg arg asp val trp leu arg
121/41                                     151/51
cgc ggc gac ggc gac ggc gac ggc gac gtt gct gcc gtt cga gga ggc gcc gga gat gac
arg gly asp gly asp gly asp gly asp val ala ala val arg gly gly ala gly asp asp
181/61                                     211/71
cag cgc ggc tgg gct cct cga gca ggc cgc cgc ggt cga gga ggc ctc cga cac cgc cgc
gln arg gly trp ala pro arg ala gly arg arg gly arg gly leu arg his arg arg
241/81                                     271/91
ggc gaa cca gtt gat gaa caa tgt gcc cca ggc gct gca aca gct ggc cca gcc cac gca
gly glu pro val asp glu gln cys ala pro gly ala ala thr ala gly pro ala his ala
301/101                                    331/111
ggg cac cac gcc ttc ttc caa gct ggg tgg cct gtg gaa gac ggt ctc gcc gca tcg gtc
gly his his ala phe phe gln ala gly trp pro val glu asp gly leu ala ala ser val
361/121                                    391/131
gcc gat cag caa cat ggt gtc gat ggc caa caa cca cat gtc gat gac caa ctc ggg tgt
ala asp gln gln his gly val asp gly gln gln pro his val asp asp gln leu gly cys
421/141                                    451/151
gtc gat gac caa cac ctt gag ctc gat gtt gaa ggg ctt tgc tcc ggc ggc ggc cgc cca
val asp asp gln his leu glu leu asp val glu gly leu cys ser gly gly gly arg pro
481/161                                    511/171
ggc cgt gca aac cgc ggc gca aaa cgg ggt ccg ggc gat gag ctc gct ggg cag ctc gct
gly arg ala asn arg gly ala lys arg gly pro gly asp glu leu ala gly gln leu ala
541/181                                    571/191
ggg ttc ttc ggg tct ggg cgg tgg ggt ggc cgc caa ctt ggg tcg ggc ggc ctc ggt cgg
gly phe phe gly ser gly arg trp gly gly arg gln leu gly ser gly gly leu gly arg
601/201                                    631/211
ttc gtt gtc ggt gcc gca ggc ctg ggc cgc ggc caa cca ggc agt cac ccc ggc ggc gcg
phe val val gly ala ala gly leu gly arg gly gln pro gly ser his pro gly gly ala
661/221                                    691/231
ggc gct gcc gct gac cag cct gac cag cgc cgc gga aag agg gcc cgg gca gat gct ggg
gly ala ala ala asp gln pro asp gln arg arg gly lys arg ala arg ala asp ala gly
721/241                                    751/251
cgg gct gcc ggt ggg gca gat ggg cgc cag ggc cgg tgg tgg gct cag tgg tgt gct gcg
arg ala ala gly gly ala asp gly arg gln gly arg trp trp ala gln trp cys ala ala
781/261                                    811/271
tgt tcc gcc gcg acc cta tgt gat gcc gca ttc tcc ggc ggc cgg cta gga gag ggg gcg
cys ser ala ala thr leu cys asp ala ala phe ser gly gly arg leu gly glu gly ala
841/281
cag act gtc gtt att tga
gln thr val val ile OPA

```

SEQ ID N° 5F

FIGURE 5F

17/185

séquence Rv1196 prédite par Cole et al. (Nature 393:537-544) et pouvant coder pour une ORF en fusion avec Seq5A'

```

1/1
atg gtg gat ttc ggg gcg tta cca ccg gag atc aac tcc gcg agg atg tac gcc ggc ccg
Met val asp phe gly ala leu pro pro glu ile asn ser ala arg met tyr ala gly pro
61/21
ggg tgc gcc tcg ctg gtg gcc gcg gct cag atg tgg gac agc gtg gcg agt gac ctg ttt
gly ser ala ser leu val ala ala ala gln met trp asp ser val ala ser asp leu phe
121/41
tcg gcc gcg tcg gcg ttt cag tcg gtg gtc tgg ggt ctg acg gtg ggg tcg tgg ata ggt
ser ala ala ser ala phe gln ser val val trp gly leu thr val gly ser trp ile gly
181/61
tcg tcg gcg ggt ctg atg gtg gcg gcg gcc tcg ccg tat gtg gcg tgg atg agc gtc acc
ser ser ala gly leu met val ala
ala ala ser pro tyr val ala trp met ser val thr
241/81
gcg ggg cag gcc gag ctg acc gcc gcc cag gtc ccg gtt gct gcg gcg gcc tac gag acg
ala gly gln ala glu leu thr ala ala gln val arg val ala ala ala tyr glu thr
301/101
gcg tat ggg ctg acg gtg ccc ccg ccg gtg atc gcc gag aac cgt gct gaa ctg atg att
ala tyr gly leu thr val pro pro pro val ile ala glu asn arg ala glu leu met ile
361/121
ctg ata gcg acc aac ctc ttg ggg caa aac acc ccg gcg atc gcg gtc aac gag gcc gaa
leu ile ala thr asn leu leu gly gln asn thr pro ala ile ala val asn glu ala glu
421/141
tac gcc gag atg tgg gcc caa gac gcc gcc gcg atg ttt ggc tac gcc gcg gcg acg gcg
tyr gly glu met trp ala gln asp ala ala ala met phe gly tyr ala ala ala thr ala
481/161
acg gcg acg gcg acg ttg ctg ccg ttc gag gag gcg ccg gag atg acc agc gcg ggt ggg
thr ala thr ala thr leu leu pro phe glu glu ala pro glu met thr ser ala gly gly
541/181
ctc ctc gag cag gcc gcc gcg gtc gag gag gcc tcc gac acc gcc gcg gcg aac cag ttg
leu leu glu gln ala ala ala val glu glu ala ser asp thr ala ala ala asn gln leu
601/201
atg aac aat gtg ccc cag gcg ctg caa cag ctg gcc cag ccc acg cag ggc acc acg cct
met asn asn val pro gln ala leu gln gln leu ala gln pro thr gln gly thr thr pro
661/221
tct tcc aag ctg ggt ggc ctg tgg aag acg gtc tcg ccg cat cgg tcg ccg atc agc aac
ser ser lys leu gly gly leu trp lys thr val ser pro his arg ser pro ile ser asn
721/241
atg gtg tcg atg gcc aac aac cac atg tcg atg acc aac tcg ggt gtg tcg atg acc aac
met val ser met ala asn asn his met ser met thr asn ser gly val ser met thr asn
781/261
acc ttg agc tcg atg ttg aag ggc ttt gct ccg gcg gcg gcc gcc cag gcc gtg caa acc
thr leu ser ser met leu lys gly phe ala pro ala ala ala ala gln ala val gln thr
841/281
gcg gcg caa aac ggg gtc ccg gcg atg agc tcg ctg ggc agc tcg ctg ggt tct tcg ggt
ala ala gln asn gly val arg ala met ser ser leu gly ser ser leu gly ser ser gly
901/301
ctg ggc ggt ggg gtg gcc gcc aac ttg ggt cgg gcg gcc tcg gtc ggt tcg ttg tcg gtg
leu gly gly gly val ala ala asn leu gly arg ala ala ser val gly ser leu ser val
961/321
ccg cag gcc tgg gcc gcg gcc aac cag gca gtc acc ccg gcg gcg ccg gcg ctg ccg ctg
pro gln ala trp ala ala ala asn gln ala val thr pro ala ala arg ala leu pro leu
1021/341
acc agc ctg acc agc gcc gcg gaa aga ggg ccc ggg cag atg ctg ggc ggg ctg ccg gtg
thr ser leu thr ser ala ala glu arg gly pro gly gln met leu gly gly leu pro val
1081/361
ggg cag atg ggc gcc agg gcc ggt ggt ggg ctc agt ggt gtg ctg cgt gtt ccg ccg cga
gly gln met gly ala arg ala gly gly gly leu ser gly val leu arg val pro pro arg
1141/381
ccc tat gtg atg ccg cat tct ccg gcg gcc ggc tag
pro tyr val met pro his ser pro ala ala gly AMB

```

SEQ ID N° 5R

FIGURE 5R
FEUILLE DE REMPLACEMENT (REGLE 26)

18/185

Seq 5P: ORF d'après Cole et al. (Nature 393:537-544) et contenant la séquence Rv1196

```

1/1                               31/11
tag gga cac gta atg gtg gat ttc ggg gcg tta cca ccg gag atc aac tcc gcg agg atg
AMB gly his val met val asp phe gly ala leu pro pro glu ile asn ser ala arg met
61/21                               91/31
tac gcc ggc ccg ggt tcg gcc tcg ctg gtg gcc gcg gct cag atg tgg gac agc gtg gcg
tyr ala gly pro gly ser ala ser leu val ala ala ala gln met trp asp ser val ala
121/41                               151/51
agt gac ctg ttt tcg gcc gcg tcg gcg ttt cag tcg gtg gtc tgg ggt ctg acg gtg ggg
ser asp leu phe ser ala ala ser ala phe gln ser val val trp gly leu thr val gly
181/61                               211/71
tcg tgg ata ggt tcg tcg gcg ggt ctg atg gtg gcg gcg gcc tcg ccg tat gtg gcg tgg
ser trp ile gly ser ser ala gly leu met val ala ala ala ser pro tyr val ala trp
241/81                               271/91
atg agc gtc acc gcg ggg cag gcc gag ctg acc gcc gcc cag gtc ccg gtt gct gcg gcg
met ser val thr ala gly gln ala glu leu thr ala ala gln val arg val ala ala ala
301/101                               331/111
gcc tac gag acg gcg tat ggg ctg acg gtg ccc ccg ccg gtg atc gcc gag aac cgt gct
ala tyr glu thr ala tyr gly leu thr val pro pro pro val ile ala glu asn arg ala
361/121                               391/131
gaa ctg atg att ctg ata gcg acc aac ctc ttg ggg caa aac acc ccg gcg atc gcg gtc
glu leu met ile leu ile ala thr asn leu leu gly gln asn thr pro ala ile ala val
421/141                               451/151
aac gag gcc gaa tac ggc gag atg tgg gcc caa gac gcc gcc gcg atg ttt ggc tac gcc
asn glu ala glu tyr gly glu met trp ala gln asp ala ala ala met phe gly tyr ala
481/161                               511/171
gcg gcg acg gcg acg gcg acg gcg acg ttg ctg ccg ttc gag gag gcg ccg gag atg acc
ala ala thr ala thr ala thr ala thr leu leu pro phe glu glu ala pro glu met thr
541/181                               571/191
agc gcg ggt ggg ctc ctc gag cag gcc gcc gcg gtc gag gag gcc tcc gac acc gcc gcg
ser ala gly gly leu leu glu gln ala ala ala val glu glu ala ser asp thr ala ala
601/201                               631/211
gcg aac cag ttg atg aac aat gtg ccc cag gcg ctg caa cag ctg gcc cag ccc acg cag
ala asn gln leu met asn asn val pro gln ala leu gln gln leu ala gln pro thr gln
661/221                               691/231
ggc acc acg cct tct tcc aag ctg ggt ggc ctg tgg aag acg gtc tcg ccg cat cgg tcg
gly thr thr pro ser ser lys leu gly gly leu trp lys thr val ser pro his arg ser
721/241                               751/251
ccg atc agc aac atg gtg tcg atg gcc aac aac cac atg tcg atg acc aac tcg ggt gtg
pro ile ser asn met val ser met ala asn asn his met ser met thr asn ser gly val
781/261                               811/271
tcg atg acc aac acc ttg agc tcg atg ttg aag ggc ttt gct ccg gcg gcg gcc gcc cag
ser met thr asn thr leu ser ser met leu lys gly phe ala pro ala ala ala gln
841/281                               871/291
gcc gtg caa acc gcg gcg caa aac ggg gtc ccg gcg atg agc tcg ctg ggc agc tcg ctg
ala val gln thr ala ala gln asn gly val arg ala met ser ser leu gly ser ser leu
901/301                               931/311
ggt tct tcg ggt ctg ggc ggt ggg gtg gcc gcc aac ttg ggt cgg gcg gcc tcg gtc ggt
gly ser ser gly leu gly gly gly val ala ala asn leu gly arg ala ala ser val gly
961/321                               991/331
tcg ttg tcg gtg ccg cag gcc tgg gcc gcg gcc aac cag gca gtc acc ccg gcg gcg ccg
ser leu ser val pro gln ala trp ala ala ala asn gln ala val thr pro ala ala arg
1021/341                               1051/351
gcg ctg ccg ctg acc agc ctg acc agc gcc gcg gaa aga ggg ccc ggg cag atg ctg ggc
ala leu pro leu thr ser leu thr ser ala ala glu arg gly pro gly gln met leu gly
1081/361 1111/371
ggg ctg ccg gtg ggg cag atg ggc gcc agg gcc ggt ggt ggg ctc agt ggt gtg ctg cgt
gly leu pro val gly gln met gly ala arg ala gly gly gly leu ser gly val leu arg
1141/381                               1171/391
gtt ccg ccg cga ccc tat gtg atg ccg cat tct ccg gcg gcc ggc tag
val pro pro arg pro tyr val met pro his ser pro ala ala gly AMB

```

SEQ ID N° 5P

FIGURE 5P
FEUILLE DE REMPLACEMENT (REGLE 26)

19/185

31/11
 GGA TCC TGA TGC AAG TGG TCC GGG ATT TGT CGG CAG CCA CCG CGG TCC CGT CGA CCA ACG
 gly ser OPA cys lys trp ser gly ile cys arg gln pro arg arg ser arg arg pro thr
 61/21
 TTG GTG CAT CCG GGC TGC GAG CAT GCA CGC ACC GAC CAG CGC GGC GAG CGC GGC TAG CTG
 leu val his pro gly cys glu his ala arg thr asp gln arg gly glu arg gly AMB leu
 121/41
 CTT GCC CAC TGT TCC TCC CTG CCG GCA CCA TGT GCG ACA AGC TTA AGC GCA GCA GTA CCG
 leu ala his cys ser ser leu pro ala pro cys ala thr ser leu ser ala ala val pro
 181/61
 GCG GTG CCT GGG CAT CCA GCA AAA CGG GGA GCT CAA GAA CGA TTC ATG AAC GAG GGG TCG
 ala val pro gly his pro ala lys arg gly ala gln glu arg phe met asn glu gly sec
 241/81
 TCA CCA ACG TCG AAA CCG ACG GTT GCC AGC CGG CCC ACG ATA TTG CGT GCT CGA GGG TCC
 ser pro thr ser lys pro thr val ala ser arg pro thr ile leu arg ala arg gly ser
 301/101
 GCT GTA CCC TCA CCG AAC GTG AGT CCC ACA CCG CGG AGG CGG GCG ACT CTG GCG TCG TTA
 ala val pro ser pro asn val ser pro thr pro arg arg arg ala thr leu ala ser leu
 361/121
 GCA GCC GAG CTC AAG GTG TCC CGC ACC ACT GTC TCG AAT GCT TTT AAC CGA CCG GAT CCA
 ala ala glu leu lys val ser arg thr thr val ser asn ala phe asn arg pro asp pro
 421/141
 GAA GGA GAA GAT C
 glu gly glu asp

SEQ ID N° 6A

FIGURE 6A

32/11
 GAT CCT GAT GCA AGT GGT CCG GGA TTT GTC GGC AGC CAC GGC GGT CCC GTC GAC CAA CGT
 asp pro asp ala ser gly pro gly phe val gly ser his gly gly pro val asp gln arg
 62/21
 TGG TGC ATC CGG GCT GCG AGC ATG CAC GCA CCG ACC AGC GCG GCG AGC GCG GCT AGC TGC
 trp cys ile arg ala ala ser met his ala pro thr ser ala ala ser ala ala ser cys
 122/41
 TTG CCC ACT GTT CCT CCC TGC CGG CAC CAT GTG CGA CAA GCT TAA GCG CAG CAG TAC CGG
 leu pro thr val pro pro cys arg his his val arg gln ala OCH ala gln gln tyr arg
 182/61
 CGG TGC CTG GGC ATC CAG CAA AAC GGG GAG CTC AAG AAC GAT TCA TGA ACG AGG GGT CGT
 arg cys leu gly ile gln gln asn gly glu leu lys asn asp ser OPA thr arg gly arg
 242/81
 CAC CAA CGT CGA AAC CGA CCG TTG CCA GCC GGC CCA CGA TAT TGC GTG CTC GAG GGT CCG
 his gln arg arg asn arg arg leu pro ala gly pro arg tyr cys val leu glu gly pro
 302/101
 CTG TAC CCT CAC CGA ACG TGA GTC CCA CAC CGC GGA GGC GGG CGA CTC TGG CGT CGT TAG
 leu tyr pro his arg thr OPA val pro his arg gly gly gly arg leu trp arg arg AMB
 362/121
 CAG CCG AGC TCA AGG TGT CCC GCA CCA CTG TCT CGA ATG CTT TTA ACC GAC CGG ATC CAG
 gln pro ser ser arg cys pro ala pro leu ser arg met leu leu thr asp arg ile gln
 422/141
 AAG GAG AAG ATC
 lys glu lys ile

SEQ ID N° 6B

FIGURE 6B
 FEUILLE DE REMPLACEMENT (REGLE 26)

20/185

33/11
 ATC CTG ATG CAA GTG GTC CGG GAT TTG TCG GCA GCC ACG GCG GTC CCG TCG ACC AAC GTT
 ile leu met gln val val arg asp leu ser ala ala thr ala val pro ser thr asn val
 63/21
 GGT GCA TCC GGG CTG CGA GCA TGC ACG CAC CGA CCA GCG CCG CGA GCG CCG CTA GCT GCT
 gly ala ser gly leu arg ala cys thr his arg pro ala arg arg ala arg leu ala ala
 123/41
 TGC CCA CTG TTC CTC CCT GCC GGC ACC ATG TGC GAC AAG CTT AAG CGC AGC AGT ACC GGC
 cys pro leu phe leu pro ala gly thr met cys asp lys leu lys arg scr ser thr gly
 183/61
 GGT GCC TGG GCA TCC AGC AAA ACG GGG AGC TCA AGA ACG ATT CAT GAA CGA GGG GTC GTC
 gly ala trp ala ser ser lys thr gly ser ser arg thr ile his glu arg gly val val
 243/81
 ACC AAC GTC GAA ACC GAC GGT TGC CAG CCG GCC CAC GAT ATT GCG TGC TCG AGG GTC CGC
 thr asn val glu thr asp gly cys gln pro ala his asp ile ala cys ser arg val arg
 303/101
 TGT ACC CTC ACC GAA CGT GAG TCC CAC ACC GCG GAG GCG GGC GAC TCT GGC GTC GTT AGC
 cys thr leu thr glu arg glu ser his thr ala glu ala gly asp ser gly val val ser
 363/121
 AGC CGA GCT CAA GGT GTC CCG CAC CAC TGT CTC GAA TGC TTT TAA CCG ACC GGA TCC AGA
 ser arg ala gln gly val pro his his cys leu glu cys phe OCH pro thr gly ser arg
 423/141
 AGG AGA AGA TC
 arg arg arg

SEQ ID N° 6C

FIGURE 6C

31/11
 CCG TCG GCA ACT TGG CCG CTG AGG TCG GCT TGA TCC CTG GGC CGA GGC GGG TCA GCC AAT
 pro ser ala thr trp pro leu arg ser ala OPA ser leu gly arg gly gly ser ala asn
 61/21
 AGC GGC TCC ATC GGC TTT GCT GGT AGC GGT TCG GCG GGA AGC TAG CCG CGA CGT TGT CGG
 ser gly ser ile gly phe ala gly ser gly ser ala gly ser AMB arg arg arg cys arg
 121/41
 TGG CCG GTG ATA TAT TCG GTC AGA CCG GTA TGG CCG CCG CTG AGG TGA TCT GCG ACA CGC
 trp pro val ile tyr trp val arg arg val trp arg arg leu arg OPA ser ala thr arg
 181/61
 CGC CGC GGT GCT CGA GCC AGG CTT ACG ACC AGG GAA TTT CGA AAA TGT TAT TCA GAA CAT
 arg arg gly ala arg ala arg leu thr thr arg glu phe arg lys cys tyr ser glu his
 241/81
 CTT GTA TCT CTT CTC CGT GCC ACC CCC TAG GTG TAG TGT TTT CGA GTA CCG GCA GAT CCC
 leu val ser leu leu arg ala thr pro AMB val AMB cys phe arg val pro ala asp pro
 301/101
 AGG TTC ACC AGG TCT CAC CAG ATC
 arg phe thr arg ser his gln ile

SEQ ID N° 7A

FIGURE 7A

FEUILLE DE REMPLACEMENT (REGLE 26)

21/185

32/11
 CGT CGG CAA CTT GGC CGC TGA GGT CGG CTT GAT CCC TGG GCC GAG GCG GGT CAG CCA ATA
 arg arg gln leu gly arg OPA gly arg leu asp pro trp ala glu ala gly gln pro ile
 62/21
 GCG GCT CCA TCG GCT TTG CTG GTA GCG GTT CGG CGG GAA GCT AGC GGC GAC GTT GTC GGT
 ala ala pro ser ala leu leu val ala val arg arg glu ala ser gly asp val val gly
 122/41
 GGC CGG TGA TAT ATT GGG TCA GAC GGG TAT GGC GGC GGC TGA GGT GAT CTG CGA CAC GCC
 gly arg OPA tyr ile gly ser asp gly tyr gly gly gly OPA gly asp leu arg his ala
 182/61
 GCC GCG GTG CTC GAG CCA GGC TTA CGA CCA GGG AAT TTC GAA AAT GTT ATT CAG AAC ATC
 ala ala val leu glu pro gly leu arg pro gly asn phe. glu asn val ile gln asn ile
 242/81
 TTG TAT CTC TTC TCC GTG CCA CCC CCT AGG TGT AGT GTT TTC GAG TAC CGG CAG ATC CCA
 leu tyr leu phe ser val pro pro pro arg cys ser val phe glu tyr arg gln ile pro
 302/101
 GGT TCA CCA GGT CTC ACC AGA TC
 gly ser pro gly leu thr arg

SEQ ID N° 7B

FIGURE 7B

33/11
 GTC GGC AAC TTG GCC GCT GAG GTC GGC TTG ATC CCT GGG CCG AGG CGG GTC AGC CAA TAG
 val gly asn leu ala ala glu val gly leu ile pro gly pro arg arg val ser gln AMB
 63/21
 CGG CTC CAT CGG CTT TGC TGG TAG CGG TTC GGC GGG AAG CTA GCG GCG ACG TTG TCG GTG
 arg leu his arg leu cys trp AMB arg phe gly gly lys leu ala ala thr leu ser val
 123/41
 GCC GGT GAT ATA TTG GGT CAG ACG GGT ATG GCG GCG GCT GAG GTG ATC TGC GAC ACG CCG
 ala gly asp ile leu gly gln thr gly met ala ala ala glu val ile cys asp thr pro
 183/61
 CCG CGG TGC TCG AGC CAG GCT TAC GAC CAG GGA ATT TCG AAA ATG TTA TTC AGA ACA TCT
 pro arg cys ser ser gln ala tyr asp gln gly ile ser lys met leu phe arg thr ser
 243/81
 TGT ATC TCT TCT CCG TGC CAC CCC CTA GGT GTA GTG TTT TCG AGT ACC GGC AGA TCC CAG
 cys ile ser ser pro cys his pro leu gly val val phe ser ser thr gly arg ser gln
 303/101
 GTT CAC CAG GTC TCA CCA GAT C
 val his gln val ser pro asp

SEQ ID N° 7C

FIGURE 7C

22/185

31/11
 CTT TGC GTG ATG TCC AAT GGC GAA AAC GAC GCC TTG TCA TCG CAA TCG TCA GCA CCG GCC
 leu cys val met ser asn gly glu asn asp ala leu ser ser gln ser ser ala pro ala
 61/21
 TAG TTT TCG CGA TGA CGC TCG TTC TGA CCG GAC TTG TGA ACG GGT TTC GGG TCG AGG CCG
 AMB phe ser arg OPA arg ser phe OPA pro asp leu OPA thr gly phe gly ser arg pro
 121/41
 AGC GAA CCG TCG ATT CCA TGG GTG TCG ACG CAT TCG TGG TCA AGG CCG GCG CGG CAG GAC
 ser glu pro ser ile pro trp val ser thr his ser trp ser arg pro ala arg gln asp
 181/61
 CGT TCC TGG GTT CGA CAC CAT TCG CCC AAA TCG ACC TGC CCC AGG TTG CTC GTG CGC CTG
 arg ser trp val arg his his ser pro lys ser thr cys pro arg leu leu val arg leu
 241/81
 GCG TCT TGG CTG CCG CCC CAC TAG CGA CTG CGC CGT CGA CGA TCC GGC AGG GCA CGT CAG
 ala ser trp leu pro pro his AMB arg leu arg arg arg ser gly arg ala arg gln
 301/101
 CGC GAA ACG TCA CCG CGT TCG GGG CAC CAG AGC ACG GAC CCG GCA TGC CGC GGG TCT CGG
 arg glu thr ser pro arg ser gly his gln ser thr asp pro ala cys arg gly ser arg
 361/121
 ACG GTC GGG CGC CAT CGA CGC CGG ACG AGG TCG CGG TGT CGA GCA CGC TGG GCC GAA ACC
 thr val gly arg his arg arg arg thr arg ser arg cys arg ala arg trp ala glu thr
 421/141
 TCG GCG ACG ATC
 ser ala thr ile

SEQ ID N° 8A

FIGURE 8A

32/11
 TTT GCG TGA TGT CCA ATG GCG AAA ACG ACG CCT TGT CAT CGC AAT CGT CAG CAC CGG CCT
 phe ala OPA cys pro met ala lys thr thr pro cys his arg asn arg gln his arg pro
 62/21
 AGT TTT CGC GAT GAC GCT CGT TCT GAC CGG ACT TGT GAA CGG GTT TCG GGT CGA GGC CGA
 ser phe arg asp asp ala arg ser asp arg thr cys glu arg val ser gly arg gly arg
 122/41
 GCG AAC CGT CGA TTC CAT GGG TGT CGA CGC ATT CGT GGT CAA GGC CGG CGC GGC AGG ACC
 ala asn arg arg phe his gly cys arg arg ile arg gly gln gly arg arg gly arg thr
 182/61
 GTT CCT GGG TTC GAC ACC ATT CGC CCA AAT CGA CCT GCC CCA GGT TGC TCG TGC GCC TGG
 val pro gly phe asp thr ile arg pro asn arg pro ala pro gly cys ser cys ala trp
 242/81
 CGT CTT GGC TGC CGC CCC ACT AGC GAC TGC GCC GTC GAC GAT CCG GCA GGG CAC GTC AGC
 arg leu gly cys arg pro thr ser asp cys ala val asp asp pro ala gly his val ser
 302/101
 GCG AAA CGT CAC CGC GTT CGG GGC ACC AGA GCA CGG ACC CGG CAT GCC GCG GGT CTC GGA
 ala lys arg his arg val arg gly thr arg ala arg thr arg his ala ala gly leu gly
 362/121
 CGG TCG GGC GCC ATC GAC GCC GGA CGA GGT CGC GGT GTC GAG CAC GCT GGG CCG AAA CCT
 arg ser gly ala ile asp ala gly arg gly arg gly val glu his ala gly pro lys pro
 422/141
 CGG CGA CGA TC
 arg arg arg

SEQ ID N° 8B

FEUILLE DE REMPLACEMENT (REGLE 26)

23/185

33/11
 TTG CGT GAT GTC CAA TGG CGA AAA CGA CGC CTT GTC ATC GCA ATC GTC AGC ACC GGC CTA
 leu arg asp val gln trp arg lys arg arg leu val ile ala ile val ser thr gly leu
 63/21 93/31
 GTT TTC GCG ATG ACG CTC GTT CTG ACC GGA CTT GTG AAC GGG TTT CGG GTC GAG GCC GAG
 val phe ala met thr leu val leu thr gly leu val asn gly phe arg val glu ala glu
 123/41 153/51
 CGA ACC GTC GAT TCC ATG GGT GTC GAC GCA TTC GTG GTC AAG GCC GGC GCG GCA GGA CCG
 arg thr val asp ser met gly val asp ala phe val val lys ala gly ala ala gly pro
 183/61 213/71
 TTC CTG GGT TCG ACA CCA TTC GCC CAA ATC GAC CTG CCC CAG GTT GCT CGT GCG CCT GGC
 phe leu gly ser thr pro phe ala gln ile asp leu pro gln val ala arg ala pro gly
 243/81 273/91
 GTC TTG GCT GCC GCC CCA CTA GCG ACT GCG CCG TCG ACG ATC CGG CAG GGC ACG TCA GCG
 val leu ala ala ala pro leu ala thr ala pro ser thr ile arg gln gly thr ser ala
 303/101 333/111
 CGA AAC GTC ACC GCG TTC GGG GCA CCA GAG CAC GGA CCC GGC ATG CCG CGG GTC TCG GAC
 arg asn val thr ala phe gly ala pro glu his gly pro gly met pro arg val ser asp
 363/121 393/131
 GGT CGG GCG CCA TCG ACG CCG GAC GAG GTC GCG GTG TCG AGC ACG CTG GGC CGA AAC CTC
 gly arg ala pro ser thr pro asp glu val ala val ser ser thr leu gly arg asn leu
 423/141
 GGC GAC GAT C
 gly asp asp

SEQ ID N° 8C

FIGURE 8C

partie de la séquence nucléotidique de seq8A

1/1 31/11
 CAG GTT GCT CGT GCG CCT GGC GTC TTG GCT GCC GCC CCA CTA GCG ACT GCG CCG TCG ACG
 gln val ala arg ala pro gly val leu ala ala ala pro leu ala thr ala pro ser thr
 61/21 91/31
 ATC CGG CAG GGC ACG TCA GCG CGA AAC GTC ACC GCG TTC GGG GCA CCA GAG CAC GGA CCC
 ile arg gln gly thr ser ala arg asn val thr ala phe gly ala pro glu his gly pro
 121/41 151/51
 GGC ATG CCG CGG GTC TCG GAC GGT CGG GCG CCA TCG ACG CCG GAC GAG GTC GCG GTG TCG
 gly met pro arg val ser asp gly arg ala pro ser thr pro asp glu val ala val ser
 181/61
 AGC ACG CTG GGC CGA AAC CTC GGC GAC GAT C
 ser thr leu gly arg asn leu gly asp asp

SEQ ID N° 8A'

FIGURE 8A'

24/185

```

1/1                               31/11
AGG TTG CTC GTG CGC CTG GCG TCT TGG CTG CCG CCC CAC TAG CGA CTG CGC CGT CGA CGA
arg leu leu val arg leu ala ser trp leu pro pro his AMB arg leu arg arg arg arg
61/21                               91/31
TCC GGC AGG GCA CGT CAG CGC GAA ACG TCA CCG CGT TCG GGG CAC CAG AGC ACG GAC CCG
ser gly arg ala arg gln arg glu thr ser pro arg ser gly his gln ser thr asp pro
121/41                               151/51
GCA TGC CGC GGG TCT CGG ACG GTC GGG CGC CAT CGA CGC CGG ACG AGG TCG CGG TGT CGA
ala cys arg gly ser arg thr val gly arg his arg arg arg thr arg ser arg cys arg
181/61
GCA CGC TGG GCC GAA ACC TCG GCG ACG ATC
ala arg trp ala glu thr ser ala thr ile

```

SEQ ID N° 8B'

FIGURE 8B'

Seq8C

```

1/1                               31/11
CCA GGT TGC TCG TGC GCC TGG CGT CTT GGC TGC CGC CCC ACT AGC GAC TGC GCC GTC GAC
pro gly cys ser cys ala trp arg leu gly cys arg pro thr ser asp cys ala val asp
61/21                               91/31
GAT CCG GCA GGG CAC GTC AGC GCG AAA CGT CAC CGC GTT CGG GGC ACC AGA GCA CGG ACC
asp pro ala gly his val ser ala lys arg his arg val arg gly thr arg ala arg thr
121/41                               151/51
CGG CAT GCC GCG GGT CTC GGA CGG TCG GGC GCC ATC GAC GCC GGA CGA GGT CGC GGT GTC
arg his ala ala gly leu gly arg ser gly ala ile asp ala gly arg gly arg gly val
181/61
GAG CAC GCT GGG CCG AAA CCT CGG CGA CGA TC
glu his ala gly pro lys pro arg arg arg

```

SEQ ID N° 8C'

FIGURE 8C'

25/185

séquence Rv2563 prédite par Cole et al. (Nature 393:537-544) et contenant seq8A'

atg

met

121/41

151/51

ctt ttt gcg gct ttg cgt gat gtc caa tgg cga aaa cga cgc ctt gtc atc gca atc gtc

leu phe ala ala leu arg asp val gln trp arg lys arg arg leu val ile ala ile val

181/61

211/71

agc acc ggc cta gtt ttc gcg atg acg ctc gtt ctg acc gga ctt gtg aac ggg ttt cgg

ser thr gly leu val phe ala met thr leu val leu thr gly leu val asn gly phe arg

241/81

271/91

gtc gag gcc gag cga acc gtc gat tcc atg ggt gtc gac gca ttc gtg gtc aag gcc ggc

val glu ala glu arg thr val asp ser met gly val asp ala phe val val lys ala gly

301/101

331/111

gcg gca gga ccg ttc ctg ggt tcg aca cca ttc gcc caa atc gac ctg ccc cag gtt gct

ala ala gly pro phe leu gly ser thr pro phe ala gln ile asp leu pro gln val ala

361/121

391/131

cgt gcg cct ggc gtc ttg gct gcc gcc cca cta gcg act gcg ccg tcg acg atc cgg cag

arg ala pro gly val leu ala ala ala pro leu ala thr ala pro ser thr ile arg gln

421/141

451/151

ggc acg tca gcg cga aac gtc acc gcg ttc ggg gca cca gag cac gga ccc ggc atg ccg

gly thr ser ala arg asn val thr ala phe gly ala pro glu his gly pro gly met pro

481/161

511/171

cgg gtc tcg gac ggt cgg gcg cca tcg acg ccg gac gag gtc gcg gtg tcg agc acg ctg

arg val ser asp gly arg ala pro ser thr pro asp glu val ala val ser ser thr leu

541/181

571/191

ggc cga aac ctc ggc gac gat ctg caa gtg ggt gcg cgc act ttg cgg atc gtc ggc atc

gly arg asn leu gly asp asp leu gln val gly ala arg thr leu arg ile val gly ile

601/201

631/211

gtg ccc gag tca acc gcg ctg gca aag att ccc aac atc ttc ctg acc acc gaa ggc cta

val pro glu ser thr ala leu ala lys ile pro asn ile phe leu thr thr glu gly leu

661/221

691/231

cag cag ttg gca tac aac gga cag ccg aca atc agt tcg atc ggg atc gac ggg atg ccc

gln gln leu ala tyr asn gly gln pro thr ile ser ser ile gly ile asp gly met pro

721/241

751/251

cga cag ctc ccg gac ggc tat cag acc gtc aat cga gcg gat gct gtc agc gat ctg atg

arg gln leu pro asp gly tyr gln thr val asn arg ala asp ala val ser asp leu met

781/261

811/271

cgc ccg ttg aag gtc gcg gtg gat gcg atc acg gtt gtg gcg gtc ttg ctg tgg atc gtt

arg pro leu lys val ala val asp ala ile thr val val ala val leu leu trp ile val

841/281

871/291

gcg gcg ttg atc gtc ggc tcg gtg gtc tac ctc tct gcg ttg gag cgg ctg cgt gac ttt

ala ala leu ile val gly ser val val tyr leu ser ala leu glu arg leu arg asp phe

901/301

931/311

gcg gtg ttc aag gcg atc ggc gtg ccg acg cgc tcg att ctg gcc ggg ctg gcg ctg cag

ala val phe lys ala ile gly val pro thr arg ser ile leu ala gly leu ala leu gln

961/321

991/331

gcg gtc gtc gtc gcg ctg ctc gcg gcg gtg gtt ggc ggc atc ctt tcg ctg ctg ttg gcg

ala val val val ala leu leu ala ala val val gly gly ile leu ser leu leu leu ala

1021/341

1051/351

ccg ttg ttc ccg atg act gtc gtg gta ccc ctg agt gcc ttc gtg gcg cta ccg gcg atc

pro leu phe pro met thr val val val pro leu ser ala phe val ala leu pro ala ile

1081/361

1111/371

gcg act gtg atc ggt ctg ctg gcc agc gtc gca gga ctg cgg cgc gtg gtg gcg atc gat

ala thr val ile gly leu leu ala ser val ala gly leu arg arg val val ala ile asp

1141/381

ccg gca cta gcg ttc gga ggt ccc tag

pro ala leu ala phe gly gly pro AMB

SEQ ID N° 8D

FEUILLE DE REMPLACEMENT (REGLE 26)

26/185

ORF prédite par_Cole et al. (Nature 393:537-544) et contenant Rv2563

```

1/1
tag gtt tca aga agg cct gtg cag gtt tcc gca gcc tgg gcc gcg gcg cca ccg aag agc
AMB val ser arg arg pro val gln val ser ala ala trp ala ala ala pro pro lys ser
61/21
ccg ccg aaa tgg gct aat cgg gtt cgc ttg gct cga tcg ccg atg atc tcg acc gcc acg
pro pro lys trp ala asn arg val arg leu ala arg ser pro met ile ser thr ala thr
121/41
acc gac ccc ctc acc tcg gtc gaa cct cgg cga acc aac gcg gca acg cca gcc cat gat
thr asp pro leu thr ser val glu pro arg arg thr asn ala ala thr pro ala his asp
181/61
cat ttg att ggg tcc acg gaa gca ggt agc ttc cgt cgc atg ctt ttt gcg gct ttg cgt
his leu ile gly ser thr glu ala gly ser phe arg arg met leu phe ala ala leu arg
241/81
gat gtc caa tgg cga aaa cga cgc ctt gtc atc gca atc gtc agc acc ggc cta gtt ttc
asp val gln trp arg lys arg arg leu val ile ala ile val ser thr gly leu val phe
301/101
gcg atg acg ctc gtt ctg acc gga ctt gtg aac ggg ttt cgg gtc gag gcc gag cga acc
ala met thr leu val leu thr gly leu val asn gly phe arg val glu ala glu arg thr
361/121
gtc gat tcc atg ggt gtc gac gca ttc gtg gtc aag gcc ggc gcg gca gga ccg ttc ctg
val asp ser met gly val asp ala phe val val lys ala gly ala ala gly pro phe leu
421/141
ggg tgc aca cca ttc gcc caa atc gac ctg ccc cag gtt gct cgt gcg cct ggc gtc ttg
gly ser thr pro phe ala gln ile asp leu pro gln val ala arg ala pro gly val leu
481/161
gct gcc gcc cca cta gcg act gcg ccg tcg acg atc cgg cag ggc acg tca gcg cga aac
ala ala ala pro leu ala thr ala pro ser thr ile arg gln gly thr ser ala arg asn
541/181
gtc acc gcg ttc ggg gca cca gag cac gga ccc gcc atg ccg ccg gtc tcg gac ggt ccg
val thr ala phe gly ala pro glu his gly pro gly met pro arg val ser asp gly arg
601/201
gcg cca tcg acg ccg gac gag gtc gcg gtg tcg agc acg ctg ggc cga aac ctc ggc gac
ala pro ser thr pro asp glu val ala val ser ser thr leu gly arg asn leu gly asp
661/221
gat ctg caa gtg ggt gcg cgc act ttg cgg atc gtc ggc atc gtg ccc gag tca acc gcg
asp leu gln val gly ala arg thr leu arg ile val gly ile val pro glu ser thr ala
721/241
ctg gca aag att ccc aac atc ttc ctg acc acc gaa ggc cta cag cag ttg gca tac aac
leu ala lys ile pro asn ile phe leu thr thr glu gly leu gln gln leu ala tyr asn
781/261
gga cag ccg aca atc agt tcg atc ggg atc gac ggg atg ccc cga cag ctc ccg gac ggc
gly gln pro thr ile ser ser ile gly ile asp gly met pro arg gln leu pro asp gly
841/281
tat cag acc gtc aat cga gcg gat gct gtc agc gat ctg atg cgc ccg ttg aag gtc gcg
tyr gln thr val asn arg ala asp ala val ser asp leu met arg pro leu lys val ala
901/301
gtg gat gcg atc acg gtt gtg gcg gtc ttg ctg tgg atc gtt gcg gcg ttg atc gtc ggc
val asp ala ile thr val val ala val leu leu trp ile val ala ala leu ile val gly
961/321
tcg gtg gtc tac ctc tct gcg ttg gag ccg ctg cgt gac ttt gcg gtg ttc aag gcg atc
ser val val tyr leu ser ala leu glu arg leu arg asp phe ala val phe lys ala ile
1021/341
ggc gtg ccg acg cgc tcg att ctg gcc ggg ctg gcg ctg cag gcg gtc gtc gtc gcg ctg
gly val pro thr arg ser ile leu ala gly leu ala leu gln ala val val val ala leu
1081/361
ctc gcg gcg gtg gtt ggc gcc atc ctt tcg ctg ctg ttg gcg ccg ttg ttc ccg atg act
leu ala ala val val gly gly ile leu ser leu leu leu ala pro leu phe pro met thr
1141/381
gtc gtg gta ccc ctg agt gcc ttc gtg gcg cta ccg gcg atc gcg act gtg atc ggt ctg
val val val pro leu ser ala phe val ala leu pro ala ile ala thr val ile gly leu
1201/401
ctg gcc agc gtc gca gga ctg ccg cgc gtg gtg gcg atc gat ccg gca cta gcg ttc gga
leu ala ser val ala gly leu arg arg val val ala ile asp pro ala leu ala phe gly
1261/421
ggg ccc tag
gly pro AMB

```

SEQ ID N° 8F

FEUILLE DE REMPLACEMENT (REGLE 26)

27/185

séquence de Rv0072 prédite par par Cole et al. (Nature 393:537-544) et
présentant plus de 77% de similarité avec Seq8D'

```

1/1                               31/11
atg ctc ttc gcg gcc ctg cgt gac atg caa tgg aga aag cgc cgc ctg gtc atc acg atc
Met leu phe ala ala leu arg asp met gln trp arg lys arg arg leu val ile thr ile
61/21                               91/31
atc agc acc ggg ctg atc ttc ggg atg acg ctt gtt ttg acc gga ctc gcg aac ggc ttc
ile ser thr gly leu ile phe gly met thr leu val leu thr gly leu ala asn gly phe
121/41                              151/51
cgg gtg gag gcc cgg cac acc gtc gat tcc atg ggt gtc gat gta ttc gtc gtc aga tcc
arg val glu ala arg his thr val asp ser met gly val asp val phe val val arg ser
181/61                              211/71
ggc gct gct gga cct ttt ctg ggt tca ata ccg ttt ccc gat gtt gac ctg gcc cga gtg
gly ala ala gly pro phe leu gly ser ile pro phe pro asp val asp leu ala arg val
241/81                              271/91
gcc gct gaa ccc ggt gtc atg gcc gcg gcc ccg ttg ggc agc gtg ggg acg atc atg aaa
ala ala glu pro gly val met ala ala ala pro leu gly ser val gly thr ile met lys
301/101                             331/111
gaa ggc acg tcg acg cga aac gtc acg gtc ttc ggc gcg ccc gag cac gga cct ggc atg
glu gly thr ser thr arg asn val thr val phe gly ala pro glu his gly pro gly met
361/121                             391/131
cca cgg gtc tca gag ggt cgg tca ccg tcg aaa ccg gac gaa gtc gcg gca tcg agc acg
pro arg val ser glu gly arg ser pro ser lys pro asp glu val ala ala ser ser thr
421/141                             451/151
atg ggc cga cac ctc ggt gac act gtc gag gtc ggc gcg cgc aga ttg cgg gtc gtt ggc
met gly arg his leu gly asp thr val glu val gly ala arg arg leu arg val val gly
481/161                             511/171
att gtg ccg aat tcc acc gcg ctg gcc aag atc ccc aat gtc ttc ctc acg acc gag ggc
ile val pro asn ser thr ala leu ala lys ile pro asn val phe leu thr thr glu gly
541/181                             571/191
tta cag aaa ttg gcg tac aac ggg cag ccg aat atc acg tcc atc ggg atc ata ggt atg
leu gln lys leu ala tyr asn gly gln pro asn ile thr ser ile gly ile ile gly met
601/201                             631/211
ccc cga cag ctg ccg gag ggt tac cag act ttc gat ccg gtg ggc gct gtc aat gat ttg
pro arg gln leu pro glu gly tyr gln thr phe asp arg val gly ala val asn asp leu
661/221                             691/231
gtg cgc cca ttg aag gtc gca gtg aat tcg atc tcg atc gtg gct gtt ttg ctg tgg att
val arg pro leu lys val ala val asn ser ile ser ile val ala val leu leu trp ile
721/241                             751/251
gtg gcg gtg ctg atc gtc ggc tcg gtg gtg tac ctt tcg gct ctt gag cgg cta cgt gac
val ala val leu ile val gly ser val val tyr leu ser ala leu glu arg leu arg asp
781/261                             811/271
ttc gcg gtg ttc aag gcg att ggc acg cca acg cgc tcg att atg gcc ggg ctc gca tta
phe ala val phe lys ala ile gly thr pro thr arg ser ile met ala gly leu ala leu
841/281                             871/291
cag gcg ctg gtc att gcg ttg ctt gcg gcg gtg gtg ggc gtc gtc ctg gcg cag gtg ttg
gln ala leu val ile ala leu leu ala ala val val gly val val leu ala gln val leu
901/301                             931/311
gca cca ctg ttt ccg atg att gtc gcg gta ccc gtc ggt gct tac ctg gcg cta ccg gtg
ala pro leu phe pro met ile val ala val pro val gly ala tyr leu ala leu pro val
961/321                             991/331
gcc gcg atc gtc atc ggt ctg ttc gct agt gtt gcc gga ttg aag cgc gtg gtg acg gtc
ala ala ile val ile gly leu phe ala ser val ala gly leu lys arg val val thr val
1021/341
gat ccc gcg cag gcg ttc gga ggt ccc tag
asp pro ala gln ala phe gly gly pro AMB

```

SEQ ID N° 8G

FEUILLE DE REMPLACEMENT (REGLE 26)

28/185

Seq8H : ORF prédite par Cole et al. (Nature 393:537-544) et contenant seq8G

1/1 31/11

tag cct ctg gga atg ctc ttc gcg gcc ctg cgt gac atg caa tgg aga aag cgc cgc ctg
 AMB pro leu gly met leu phe ala ala leu arg asp met gln trp arg lys arg arg leu
 61/21 91/31

gtc atc acg atc atc agc acc ggg ctg atc ttc ggg atg acg ctt gtt ttg acc gga ctc
 val ile thr ile ile ser thr gly leu ile phe gly met thr leu val leu thr gly leu
 121/41 151/51

gcg aac ggc ttc cgg gtg gag gcc cgg cac acc gtc gat tcc atg ggt gtc gat gta ttc
 ala asn gly phe arg val glu ala arg his thr val asp ser met gly val asp val phe
 181/61 211/71

gtc gtc aga tcc ggc gct gct gga cct ttt ctg ggt tca ata ccg ttt ccc gat gtt gac
 val val arg ser gly ala ala gly pro phe leu gly ser ile pro phe pro asp val asp
 241/81 271/91

ctg gcc cga gtg gcc gct gaa ccc ggt gtc atg gcc gcg gcc ccg ttg ggc agc gtg ggg
 leu ala arg val ala ala glu pro gly val met ala ala ala pro leu gly ser val gly
 301/101 331/111

acg atc atg aaa gaa ggc acg tcg acg cga aac gtc acg gtc ttc ggc gcg ccc gag cac
 thr ile met lys glu gly thr ser thr arg asn val thr val phe gly ala pro glu his
 361/121 391/131

gga cct ggc atg cca cgg gtc tca gag ggt cgg tca ccg tcg aaa ccg gac gaa gtc gcg
 gly pro gly met pro arg val ser glu gly arg ser pro ser lys pro asp glu val ala
 421/141 451/151

gca tcg agc acg atg ggc cga cac ctc ggt gac act gtc gag gtc ggc gcg cgc aga ttg
 ala ser ser thr met gly arg his leu gly asp thr val glu val gly ala arg arg leu
 481/161 511/171

cgg gtc gtt ggc att gtg ccg aat tcc acc gcg ctg gcc aag atc ccc aat gtc ttc ctc
 arg val val gly ile val pro asn ser thr ala leu ala lys ile pro asn val phe leu
 541/181 571/191

acg acc gag ggc tta cag aaa ttg gcg tac aac ggg cag ccg aat atc acg tcc atc ggg
 thr thr glu gly leu gln lys leu ala tyr asn gly gln pro asn ile thr ser ile gly
 601/201 631/211

atc ata ggt atg ccc cga cag ctg ccg gag ggt tac cag act ttc gat cgg gtg ggc gct
 ile ile gly met pro arg gln leu pro glu gly tyr gln thr phe asp arg val gly ala
 661/221 691/231

gtc aat gat ttg gtg cgc cca ttg aag gtc gca gtg aat tcg atc tcg atc gtg gct gtt
 val asn asp leu val arg pro leu lys val ala val asn ser ile ser ile val ala val
 721/241 751/251

ttg ctg tgg att gtg gcg gtg ctg atc gtc ggc tcg gtg gtg tac ctt tcg gct ctt gag
 leu leu trp ile val ala val leu ile val gly ser val val tyr leu ser ala leu glu
 781/261 811/271

cgg cta cgt gac ttc gcg gtg ttc aag gcg att ggc acg cca acg cgc tcg att atg gcc
 arg leu arg asp phe ala val phe lys ala ile gly thr pro thr arg ser ile met ala
 841/281 871/291

ggg ctc gca tta cag gcg ctg gtc att gcg ttg ctt gcg gcg gtg gtg ggc gtc gtc ctg
 gly leu ala leu gln ala leu val ile ala leu leu ala ala val val gly val val leu
 901/301 931/311

gcg cag gtg ttg gca cca ctg ttt ccg atg att gtc gcg gta ccc gtc ggt gct tac ctg
 ala gln val leu ala pro leu phe pro met ile val ala val pro val gly ala tyr leu
 961/321 991/331

gcg cta ccg gtg gcc gcg atc gtc atc ggt ctg ttc gct agt gtt gcc gga ttg aag cgc
 ala leu pro val ala ala ile val ile gly leu phe ala ser val ala gly leu lys arg
 1021/341 1051/351

gtg gtg acg gtc gat ccc gcg cag gcg ttc gga ggt ccc tag
 val val thr val asp pro ala gln ala phe gly gly pro AMB

SEQ ID N° 8H

FEUILLE DE REMPLACEMENT (REGLE 26)

29/185

31/11
 CGA GGC CGA GCG AAC CGT CGA TTC CAT GGG TGT CGA CGC ATT CGT GGT CAA GGC CGG CGC
 arg gly arg ala asn arg arg phe his gly cys arg arg ile arg qly gln gly arg arg
 61/21
 GGC AGG ACC GTT CCT GGG TTC GAC ACC ATT CGC CCA AAT CGA CCT GCC CCA GGT TGC TCG
 gly arg thr val pro gly phe asp thr ile arg pro asn arg pro ala pro gly cys ser
 121/41
 TGC GCC TGG CGT CTT GGC TGC CGC CCC ACT AGC GAC TGC GCC GTC GAC GAT CCG GCA GGG
 cys ala trp arg leu gly cys arg pro thr ser asp cys ala val asp asp pro ala gly
 181/61
 CAC GTC AGC GCG AAA CGT CAC CGC GTT CGG GGC ACC AGA GCA CGG ACC CGG CAT GCC GCG
 his val ser ala lys arg his arg val arg gly thr arg ala arg thr arg his ala ala
 241/81
 GGT CTC GGA CGG TCG GGC GCC ATC GAC GCC GGA CGA GGT CGC GGT GTC GAG CAC GCT GGG
 gly leu gly arg ser gly ala ile asp ala gly arg gly arg gly val glu his ala gly
 301/101
 CCG AAA CCT CGG CGA CGA TC
 pro lys pro arg arg arg

SEQ ID N° 9A

FIGURE 9A

32/11
 GAG GCC GAG CGA ACC GTC GAT TCC ATG GGT GTC GAC GCA TTC GTG GTC AAG GCC GGC GCG
 glu ala glu arg thr val asp ser met gly val asp ala phe val val lys ala gly ala
 62/21
 GCA GGA CCG TTC CTG GGT TCG ACA CCA TTC GCC CAA ATC GAC CTG CCC GAG GTT GCT CGT
 ala gly pro phe leu gly ser thr pro phe ala gln ile asp leu pro gln val ala arg
 122/41
 GCG CCT GGC GTC TTG GCT GCC GCC CCA CTA GCG ACT GCG CCG TCG ACG ATC CGG CAG GGC
 ala pro gly val leu ala ala ala pro leu ala thr ala pro ser thr ile arg gln gly
 182/61
 ACG TCA GCG CGA AAC GTC ACC GCG TTC GGG GCA CCA GAG CAC GGA CCC GGC ATG CCG CGG
 thr ser ala arg asn val thr ala phe gly ala pro glu his gly pro gly met pro arg
 242/81
 GTC TCG GAC GGT CGG GCG CCA TCG ACG CCG GAC GAG GTC GCG GTG TCG AGC ACG CTG GGC
 val ser asp gly arg ala pro ser thr pro asp glu val ala val ser ser thr leu gly
 302/101
 CGA AAC CTC GGC GAC GAT C
 arg asn leu gly asp asp

SEQ ID N° 9B

FIGURE 9B

30/185

33/11

AGG CCG AGC GAA CCG TCG ATT CCA TGG GTG TCG ACG CAT TCG TGG TCA AGG CCG GCG CGG
 arg pro ser glu pro ser ile pro trp val ser thr his ser trp ser arg pro ala arg
 63/21 93/31
 CAG GAC CGT TCC TGG GTT CGA CAC CAT TCG CCC AAA TCG ACC TGC CCC AGG TTG CTC GTG
 gln asp arg ser trp val arg his his ser pro lys ser thr cys pro arg leu leu val
 123/41 153/51
 CGC CTG GCG TCT TGG CTG CCG CCC CAC TAG CGA CTG CGC CGT CGA CGA TCC GGC AGG GCA
 arg leu ala ser trp leu pro pro his AMB arg leu arg arg arg arg ser gly arg ala
 183/61 213/71
 CGT CAG CGC GAA ACG TCA CCG CGT TCG GGG CAC CAG AGC ACG GAC CCG GCA TGC CGC GGG
 arg gln arg glu thr ser pro arg ser gly his gln ser thr asp pro ala cys arg gly
 243/81 273/91
 TCT CGG ACG GTC GGG CGC CAT CGA CGC CGG ACG AGG TCG CGG TGT CGA GCA CGC TGG GCC
 ser arg thr val gly arg his arg arg arg thr arg ser arg cys arg ala arg trp ala
 303/101
 GAA ACC TCG GCG ACG ATC
 glu thr.ser ala thr ile

SEQ ID N° 9C

FIGURE 9C

31/11

TTA ACG ACT CAG ACG GAA ACG CTT GAA CCG CGA GGT CGC TCC GGA CAC CAA TTT GAC TCG
 leu thr thr gln thr glu thr leu glu pro arg gly arg ser gly his gln phe asp ser
 61/21 91/31
 GCT CTT TGG CAA TTG AAG GTG AGC TGC GAG CAG CCG GGT GAC CGC ATC GTT GGC CTT GCC
 ala leu trp gln leu lys val ser cys glu gln pro gly asp arg ile val gly leu ala
 121/41 151/51
 ATC AAT CGC CGG CTC GCG GAC GTA GAT AAT CAG CTC ACC GTT GGG ACC GAC CTC GAC CAG
 ile asn arg arg leu ala asp val asp asn gln leu thr val gly thr asp leu asp gln
 181/61 211/71
 GGG TCC TTT GTG ACT GCC GGG CTT GAC GCG GAC GAC CAC AGA GTC GGT CAT CGC CTA AGG
 gly ser phe val thr ala gly leu asp ala asp asp his arg val gly his arg leu arg
 241/81 271/91
 CTA CCG TTC TGA CCT GGG GCT GCG TGG GCG CCG ACG ACG TGA GGC ACG TCA TGT CTC AGC
 leu pro phe OPA pro gly ala ala trp ala pro thr thr OPA gly thr ser cys leu ser
 301/101 331/111
 GGC CCA CCG CCA CCT CGG TCG CCG GCA GTA TGT CAG CAT GTG CAG ATG ACT CCA CGC AGC
 gly pro pro pro pro arg ser pro ala val cys gln his val gln met thr pro arg ser
 361/121 391/131
 CTT GTT CGC ATC GTT GGT GTC GTG GTT GCG ACG ACC TTG GCG CTG GTG AGC GCA CCC GCC
 leu val arg ile val gly val val val ala thr thr leu ala leu val ser ala pro ala
 421/141
 GGC GGT CGT GCC GCG CAT GCG GAT C
 gly gly arg ala ala his ala asp

SEQ ID N° 10A

FIGURE 10A

FEUILLE DE REMPLACEMENT (REGLE 26)

31/185

32/11
TAA CGA CTC AGA CGG AAA CGC TTG AAC CGC GAG GTC GCT CCG GAC ACC AAT TTG ACT CGG
OCH arg leu arg arg lys arg leu asn arg glu val ala pro asp thr asn leu thr arg
62/21
CTC TTT GGC AAT TGA AGG TGA GCT GCG AGC AGC CGG GTG ACC GCA TCG TTG GCC TTG CCA
leu phe gly asn OPA arg OPA ala ala ser ser arg val thr ala ser leu ala leu pro
122/41
TCA ATC GCC GGC TCG CGG ACG TAG ATA ATC AGC TCA CCG TTG GGA CCG ACC TCG ACC AGG
ser ile ala gly ser arg thr AMB ile ile ser ser pro leu gly pro thr ser thr arg
182/61
GGT CCT TTG TGA CTG CCG GGC TTG ACG CGG ACC ACC ACA GAG TCG GTC ATC GCC TAA GGC
gly pro leu OPA leu pro gly leu thr arg thr thr thr glu ser val ile ala OCH gly
242/81
TAC CGT TCT GAC CTG GGG CTG CGT GGG CGC CGA CGA CGT GAG GCA CGT CAT GTC TCA GCG
tyr arg ser asp leu gly leu arg gly arg arg arg arg glu ala arg his val ser ala
302/101
GCC CAC CGC GAC CTC GGT CGC CGG CAG TAT GTC AGC ATG TGC AGA TGA CTC CAC GCA GCC
ala his arg his leu gly arg arg gln tyr val ser met cys arg OPA leu his ala ala
362/121
TTG TTC GCA TCG TTG GTG TCG TGG TTG CGA CGA CCT TGG CGC TGG TGA GCG CAC CCG CCG
leu phe ala ser leu val ser trp leu arg arg pro trp arg trp OPA ala his pro pro
422/141
GCG GTC GTG CCG CGC ATG CGG ATC
ala val val pro arg met arg Ile

SEQ ID N° 10B

FIGURE 10B

33/11
AAC GAC TCA GAC GGA AAC GCT TGA ACC GCG AGG TCG CTC CGG ACA CCA ATT TGA CTC GGC
asn asp ser asp gly asn ala OPA thr ala arg ser leu arg thr pro ile OPA leu gly
63/21
TCT TTG GCA ATT GAA GGT GAG CTG CGA GCA GCC GGG TGA CCG CAT CGT TGG CCT TGC CAT
ser leu ala ile glu gly glu leu arg ala ala gly OPA pro his arg trp pro cys his
123/41
CAA TCG CCG GCT CGC GGA CGT AGA TAA TCA GCT CAC CGT TGG GAC CGA CCT CGA CCA GGG
gln ser pro ala arg gly arg arg OCH ser ala his arg trp asp arg pro arg pro gly
183/61
GTC CTT TGT GAC TGC CGG GCT TGA CGC GGA CGA CCA CAG AGT CGG TCA TCG CCT AAG GCT
val leu cys asp cys arg ala OPA arg gly arg pro gln ser arg ser ser pro lys ala
243/81
ACC GTT CTG ACC TGG GGC TGC GTG GGC GCC GAC GAC GTG AGG CAC GTC ATG TCT CAG CGG
thr val leu thr trp gly cys val gly ala asp asp val arg his val met ser gln arg
303/101
CCC ACC GCC ACC TCG GTC GCC GGC AGT ATG TCA GCA TGT GCA GAT GAC TCC ACG CAG CCT
pro thr ala thr ser val ala gly ser met ser ala cys ala asp asp ser thr gln pro
363/121
TGT TCG CAT CGT TGG TGT CGT GGT TGC GAC GAC CTT GGC GCT GGT GAG CGC ACC CGC CGG
cys ser his arg trp cys arg gly cys asp asp leu gly ala gly glu arg thr arg arg
423/141
CGG TCG TGC CGC GCA TGC GGA TC
arg ser cys arg ala cys gly

SEQ ID N° 10C

FIGURE 10C
FEUILLE DE REMPLACEMENT (REGLE 26)

32/185

31/11
 CCC GAA GAG GTC CCC CGT TTT GTT AAT TTT TAA AAA ATT TGT GTC ACA AAC CGG GGT ACC
 pro glu glu val pro arg phe val asp phe OCH lys ile cys val thr lys arg gly thr
 61/21
 AAG GCA TAA AAC CTA GTA CCT GGG GCG GCG GAT TCA ACG AAA ACC GAG TGG GGG TAG TCA
 lys ala OCH asn leu val pro gly ala ala asp ser thr lys thr glu trp gly AMB ser
 121/41
 GGG GCG TGC ATT CCG ACG ACC CTG TAC GAC CCG CTG GTG GCA ACG CCG ATG AGT GCG CCG
 gly ala cys ile pro thr thr leu tyr asp pro leu val ala thr pro met ser ala pro
 181/61
 ACG AAG GCC GAG CGA CCG GCT GCC GGC GCT GAC CGC CGC GGA AGC CGC CGA GTG CAT GGT
 thr lys ala glu arg arg ala ala gly ala asp arg arg gly ser arg arg val asp gly
 241/81
 CAC CAC CGC CCG CAC CCG ACC GGT ACG GAT CGC GCC TCG GGT TAC CGT CGC CGT CAA CGC
 his his arg pro his pro thr gly thr asp arg ala ser gly tyr arg arg arg gln arg
 301/101
 GCT GGA CAG CAT CCG TCC CCG CTG GGT CAA TGC ACT CAT GCA GCG CCG CAA CGA ACA GCT
 ala gly gln his arg ser pro leu gly gln cys thr his ala ala pro gln arg thr ala
 361/121
 CAA CCC TTG AAC CCG GTC CCG GCC TGC CGA CCC TCG GCC GCC GGC GTG CCG CTA CGT GAT
 gln pro leu asn arg val pro ala cys arg pro ser ala ala gly val pro leu arg asp
 421/141
 AGA CAC AGG GCC ATG GAA ATC CTG GCC AGC CGG ATG CTA CTT CGG CCG GCG GAC TAT CAG
 arg his arg ala met glu ile leu ala ser arg met leu leu arg pro ala asp tyr gln
 481/161
 CGG TCG CTG AGC TTC TAC CGT GAC CAG ATC
 arg ser leu ser phe tyr arg asp gln ile

SEQ ID N° 11A

FIGURE 11A

32/11
 CCG AAG AGG TCC CCC GTT TTG TTA ATT TTT AAA AAA TTT GTG TCA CAA AGC GGG GTA CCA
 pro lys arg ser pro val leu leu ile phe lys lys phe val ser gln ser gly val pro
 62/21
 AGG CAT AAA ACC TAG TAC CTG GGG CCG CGG ATT CAA CGA AAA CCG AGT GGG GGT AGT CAG
 arg his lys thr AMB tyr leu gly arg arg ile gln arg lys pro ser gly gly ser gln
 122/41
 GGG CGT GCA TTC CGA CGA CCC TGT ACG ACC CGC TGG TGG CAA CGC CGA TGA GTG CCC CGA
 gly arg ala phe arg arg pro cys thr thr arg trp trp gin arg arg OPA val arg arg
 182/61
 CGA AGG ccg AGC GAC GGG CTG CCG GCG CTG ACC GCC GCG GAA GCC GCC GAG TGG ATG GTC
 arg arg pro ser asp gly leu pro ala leu thr ala ala glu ala ala glu trp met val
 242/81
 ACC ACC GCC CGC ACC CGA CCG GTA CCG ATC GCG CCT CGG GTT ACC GTC GCC GTC ACC GCG
 thr thr ala arg thr arg pro val arg ile ala pro arg val thr val ala val asn ala
 302/101
 CTG GAC AGC ATC GGT CCC CGC TGG GTC AAT GCA CTC ATG CAG CGC CGC AAC GAA CAG CTC
 leu asp ser ile gly pro arg trp val asn ala leu met gln arg arg asn glu gln leu
 362/121
 AAC CCT TGA ACC GGG TCC CGG CCT GCC GAC CCT CGG CCG CCG GCG TGC CGC TAC GTG ATA
 asn pro OPA thr gly ser arg pro ala asp pro arg pro pro ala cys arg tyr val ile
 422/141
 GAC ACA GGG CCA TGG AAA TCC TGG CCA GCC GGA TGC TAC TTC GCC CGG CGG ACT ATC AGC
 asp thr gly pro trp lys ser trp pro ala gly cys tyr phe gly arg arg thr ile ser
 482/161
 GGT CGC TGA GCT TCT ACC GTG ACC AGA TC
 gly arg OPA ala ser thr val thr arg

SEQ ID N° 11B

FIGURE 11B

FEUILLE DE REMPLACEMENT (REGLE 26)

33/185

33/11

CGA AGA GGT CCC CCG TTT TGT TAA TTT TTA AAA AAT TTG TGT CAC AAA GCG GGG TAC CAA
 arg arg gly pro pro phe cys OCH phe leu lys asn leu cys his lys ala gly tyr gln
 63/21 93/31

GGC ATA AAA CCT AGT ACC TGG GGC GGC GGA TTC AAC GAA AAC CGA GTG GGG GTA GTC AGG
 gly ile lys pro ser thr trp gly gly gly phe asn glu asn arg val gly val val arg
 123/41 153/51

GGC GTG CAT TCC GAC GAC CCT GTA CGA CCC GCT GGT GGC AAC GCC GAT GAG TGC GCC GAC
 gly val his ser asp asp pro val arg pro ala gly gly asn ala asp glu cys ala asp
 183/61 213/71

GAA GGC CGA GCG ACG GGC TGC CGG CGC TGA CCG CCG CGG AAG CCG CCG AGT GGA TGG TCA
 glu gly arg ala thr gly cys arg arg OPA pro pro arg lys pro pro ser gly trp ser
 243/81 273/91

CCA CCG CCC GCA CCC GAC CGG TAC GGA TCG CGC CTC GGG TTA CCG TCG CCG TCA ACG CGC
 pro pro pro ala pro asp arg tyr gly ser arg leu gly leu pro ser pro ser thr arg
 303/101 333/111

TGG ACA GCA TCG GTC CCC GCT GGG TCA ATG CAC TCA TGC AGC GCC GCA ACG AAC AGC TCA
 trp thr ala ser val pro ala gly ser met his ser cys ser ala ala thr asn ser ser
 363/121 393/131

ACC CTT GAA CCG GGT CCC GGC CTG CCG ACC CTC GGC CGC CGG CGT GCC GCT ACG TGA TAG
 thr leu glu pro gly pro gly leu pro thr leu gly arg arg arg ala ala thr OPA AMB
 423/141 453/151

ACA CAG GGC CAT GGA AAT CCT GGC CAG CCG GAT GCT ACT TCG GCC GGC GGA CTA TCA GCG
 thr gln gly his gly asn pro gly gln pro asp ala thr ser ala gly gly leu ser ala
 483/161

GTC GCT GAG CTT CTA CCG TGA CCA GAT C
 val ala glu leu leu pro OPA pro asp

SEQ ID N° 11C

FIGURE 11C

partie de la séquence nucléotidique de Seq11

1/1 31/11

CGT CGC CGT CAA CGC GCT GGA CAG CAT CGG TCC CCG CTG GGT CAA TGC ACT CAT GCA GCG
 arg arg arg gln arg ala gly gln his arg ser pro leu gly gln cys thr his ala ala
 61/21 91/31

CCG CAA CGA ACA GCT CAA CCC TTG AAC CGG GTC CCG GCC TGC CGA CCC TCG GCC GCC GGC
 pro gln arg thr ala gln pro leu asn arg val pro ala cys arg pro ser ala ala gly
 121/41 151/51

GTG CCG CTA CGT GAT AGA CAC AGG GCC ATG GAA ATC CTG GCC AGC CGG ATG CTA CTT CGG
 val pro leu arg asp arg his arg ala met glu ile leu ala ser arg met leu leu arg
 181/61 211/71

CCG GCG GAC TAT CAG CGG TCG CTG AGC TTC TAC CGT GAC CAG ATC
 pro ala asp tyr gln arg ser leu ser phe tyr arg asp gln ile

SEQ ID N° 11A'

FIGURE 11A'

34/185

```

1/1                               31/11
GTC GCC GTC AAC GCG CTG GAC AGC ATC GGT CCC CGC TGG GTC AAT GCA CTC ATG CAG CGC
val ala val asn ala leu asp ser ile gly pro arg trp val asn ala leu met gln arg
61/21                               91/31
CGC AAC GAA CAG CTC AAC CCT TGA ACC GGG TCC CGG CCT GCC GAC CCT CGG CCG CCG CGC
arg asn glu gln leu asn pro OPA thr gly ser arg pro ala asp pro arg pro pro ala
121/41                               151/51
TGC CGC TAC GTG ATA GAC ACA GGG CCA TGG AAA TCC TGG CCA GCC GGA TGC TAC TTC GGC
cys arg tyr val ile asp thr gly pro trp lys ser trp pro ala gly cys tyr phe gly
181/61                               211/71
CGG CGG ACT ATC AGC GGT CGC TGA GCT TCT ACC GTG ACC AGA TC
arg arg thr ile ser gly arg OPA ala ser thr val thr arg

```

SEQ ID N° 11B'

FIGURE 11B'

```

1/1                               31/11
TCG CCG TCA ACG CGC TGG ACA GCA TCG GTC CCC GCT GGG TCA ATG CAC TCA TGC AGC GCC
ser pro ser thr arg trp thr ala ser val pro ala gly ser met his ser cys ser ala
61/21                               91/31
GCA ACG AAC AGC TCA ACC CTT GAA CCG GGT CCC GGC CTG CCG ACC CTC GGC CGC CGG CGT
ala thr asn ser ser thr leu glu pro gly pro gly leu pro thr leu gly arg arg arg
121/41                               151/51
GCC GCT ACG TGA TAG ACA CAG GGC CAT GGA AAT CCT GGC CAG CCG GAT GCT ACT TCG GCC
ala ala thr OPA AMB thr gln gly his gly asn pro gly gln pro asp ala thr ser ala
181/61                               211/71
GGC GGA CTA TCA GCG GTC GCT GAG CTT CTA CCG TGA CCA GAT C
gly gly leu ser ala val ala glu leu leu pro OPA pro asp

```

SEQ ID N° 11C'

FIGURE 11C'

séquence Rv0546c prédite par par Cole et al. (Nature 393:537-544) et contenant Seq11A'

```

1/1                               31/11
atg gaa atc ctg gcc agc cgg atg cta ctt cgg ccg gcg gac tat cag cgg tcg ctg agc
Met glu ile leu ala ser arg met leu leu arg pro ala asp tyr gln arg ser leu ser
61/21                               91/31
ttc tac cgt gac cag atc ggg ctg gcg att gcc cgt gaa tac ggg gcc ggc aca gtg ttt
phe tyr arg asp gln ile gly leu ala ile ala arg glu tyr gly ala gly thr val phe
121/41                               151/51
ttc gcc ggt cag tca ctg ctc gaa ctg gcc ggt tac ggc gag ccg gac cat tcg cgg gga
phe ala gly gln ser leu leu glu leu ala gly tyr gly glu pro asp his ser arg gly
181/61                               211/71
cct ttt ccc ggc gcg ctg tgg ctg cag gtg cgc gac ctc gag gct acc cag acc gag ctg
pro phe pro gly ala leu trp leu gln val arg asp leu glu ala thr gln thr glu leu
241/81                               271/91
gtc agc cga ggc gtg tcg atc gct cgc gag ccc cgc cgc gaa ccg tgg ggc ctg cac gag
val ser arg gly val ser ile ala arg glu pro arg arg glu pro trp gly leu his glu
301/101                               331/111
atg cat gtg acc gac cca gac ggg atc aca ctg ata ttc gtc gag gtt ccc gag ggt cac
met his val thr asp pro asp gly ile thr leu ile phe val glu val pro glu gly his
361/121
ccg ctg cgt aca gac acc cgg gcg tga
pro leu arg thr asp thr arg ala OPA

```

SEQ ID N° 11D

FEUILLE DE REMPLACEMENT (REGLE 26)

35/185

ORF prédite par par Cole et al. (Nature 393:537-544) et contenant Rv0546c

```

1/1                                31/11
tag tca ggg cgt gca ttc gac gac gct gta cta ccc gct ggt. ggc aac tcc gat gat tgc
AMB ser gly arg ala phe asp asp ala val leu pro ala gly gly asn ser asp asp cys
61/21                                91/31
gcc gac gaa ggc cta cga cgg gct gcc ggc gct gac cgc cgc gga agc cgc cga gtg gat
ala asp glu gly leu arg arg ala ala gly ala asp arg arg gly ser arg arg val asp
121/41                                151/51
ggt cac cgc cgc ccg cac ccg acc ggt gcg gat cgc gcc tcg ggt tgc cgt cgc cgt caa
gly his arg arg pro his pro thr gly ala asp arg ala ser gly cys arg arg arg gln
181/61                                211/71
cgc gct gga cag cat cgg tcc ccg ctg ggt caa tgc act cat gca gcg ccg caa cga aca
arg ala gly gln his arg ser pro leu gly gln cys thr his ala ala pro gln arg thr
241/81                                271/91
gct caa ccc ttg aac cgg gtc ccg gcc tgc cga ccc tcg gcc gcc ggc gtg ccg cta cgt
ala gln pro leu asn arg val pro ala cys arg pro ser ala ala gly val pro leu arg
301/101                               331/111
gat aga cac agg gcc atg gaa atc ctg gcc agc cgg atg cta ctt cgg ccg gcg gac tat
asp arg his arg ala met glu ile leu ala ser arg met leu leu arg pro ala asp tyr
361/121                               391/131
cag cgg tcg ctg agc ttc tac cgt gac cag atc ggg ctg gcg att gcc cgt gaa tac ggg
gln arg ser leu ser phe tyr arg asp gln ile gly leu ala ile ala arg glu tyr gly
421/141                               451/151
gcc ggc aca gtg ttt ttc gcc ggt cag tca ctg ctc gaa ctg gcc ggt tac ggc gag ccg
ala gly thr val phe phe ala gly gln ser leu leu glu leu ala gly tyr gly glu pro
481/161                               511/171
gac cat tcg cgg gga cct ttt ccc ggc gcg ctg tgg ctg cag gtg cgc gac ctc gag gct
asp his ser arg gly pro phe pro gly ala leu trp leu gln val arg asp leu glu ala
541/181                               571/191
acc cag acc gag ctg gtc agc cga ggc gtg tcg atc gct cgc gag ccc cgc cgc gaa ccg
thr gln thr glu leu val ser arg gly val ser ile ala arg glu pro arg arg glu pro
601/201                               631/211
tgg ggc ctg cac gag atg cat gtg acc gac cca gac ggg atc aca ctg ata ttc gtc gag
trp gly leu his glu met his val thr asp pro asp gly ile thr leu ile phe val glu
661/221                               691/231
gtt ccc gag ggt cac ccg ctg cgt aca gac acc cgg gcg tga
val pro glu gly his pro leu arg thr asp thr arg ala OPA

```

SEQ ID N° 11F

FIGURE 11F

36/185

```

1/1                               31/11
gac cga agg gat ttc gcg act aac tcg gcc tgt aag gca acg cga ggt ctt cat gcc gag
asp arg arg asp phe ala thr asn ser ala cys lys ala thr arg gly leu his ala glu
61/21                               91/31
gac gta gac agg aag aga cag gga agc tga tga cgt cgc gta ccg gac cgc cat tct gtc
asp val asp arg lys arg gln gly ser OPA OPA arg arg val pro asp arg his ser val
121/41                               151/51
gag tct ttc cga gtt cag caa caa tcg aca cag aag cgg gga cca gac cgg gag gac gac
glu ser phe arg val gln gln gln ser thr gln lys arg gly pro asp arg glu asp asp
181/61                               211/71
gcg gcc cgg gcc gct tcg ggc cga gtg tct gag taa gac cag agt cac ggg tcc gtg tgt
ala ala arg ala ala ser gly arg val ser glu OCH asp gln ser his gly ser val cys
241/81                               271/91
gac aac cgc gcg gaa ttc aat cgg atg gcg ggc ggg acc gga ttg cgc cgg tca ccg agg
asp asn arg ala glu phe asn arg met ala gly gly thr gly leu arg arg ser pro arg
301/101
aac ctc cgg agt gat c
asn leu arg ser asp

```

SEQ ID N° 12A

FIGURE 12A

```

1/1                               31/11
acc gaa ggg att tcg cga cta act cgg cct gta agg caa cgc gag gtc ttc atg ccg agg
thr glu gly ile ser arg leu thr arg pro val arg gln arg glu val phe met pro arg
61/21                               91/31
acg tag aca gga aga gac agg gaa gct gat gac gtc gcg tac cgg acc gcc att ctg tcg
thr AMB thr gly arg asp arg glu ala asp asp val ala tyr arg thr ala ile leu ser
121/41                               151/51
agt ctt tcc gag ttc agc aac aat cga cac aga agc ggg gac cag acc ggg agg acg acg
ser leu ser glu phe ser asn asn arg his arg ser gly asp gln thr gly arg thr thr
181/61                               211/71
cgg ccc ggg ccg ctt cgg gcc gag tgt ctg agt aag acc aga gtc acg ggt ccg tgt gtg
arg pro gly pro leu arg ala glu cys leu ser lys thr arg val thr gly pro cys val
241/81                               271/91
aca acc gcg cgg aat tca atc gga tgg cgg gcg gga ccg gat tgc gcc ggt cac cga gga
thr thr ala arg asn ser ile gly trp arg ala gly pro asp cys ala gly his arg gly
301/101
acc tcc gga gtg atc
thr ser gly val ile

```

SEQ ID N° 12B

FIGURE 12B

37/185

```

1/1                               31/11
ccg aag gga ttt cgc gac taa ctc ggc ctg taa ggc aac gcg agg tct tca tgc cga gga
pro lys gly phe arg asp OCH leu gly leu OCH gly asn ala arg ser ser cys arg gly
61/21                               91/31
cgt aga cag gaa gag aca ggg aag ctg atg acg tcg cgt acc gga ccg cca ttc tgt cga
arg arg gln glu glu thr gly lys leu met thr ser arg thr gly pro pro phe cys arg
121/41                               151/51
gtc ttt ccg agt tca gca aca atc gac aca gaa gcg ggg acc aga ccg gga gga cga cgc
val phe pro ser ser ala thr ile asp thr glu ala gly thr arg pro gly gly arg arg
181/61                               211/71
ggc ccg ggc cgc ttc ggg ccg agt gtc tga gta aga cca gag tca ccg gtc cgt gtg tga
gly pro gly arg phe gly pro ser val OPA val arg pro glu ser arg val arg val OPA
241/81                               271/91
caa ccg cgc gga att caa tcg gat ggc ggg ccg gac ccg att gcg ccg gtc acc gag gaa
gln pro arg gly ile gln ser asp gly gly arg asp arg ile ala pro val thr glu glu
301/101
cct ccg gag tga tc
pro pro glu OPA

```

SEQ ID N° 12C

FIGURE 12C

```

1/1                               31/11
GGG ATT TCG TTG CCC GAT GGA TTG TTT GTA CGG TTT GGG AAA AAC ACT TGA AGT CCT TTT
gly ile ser leu pro asp gly leu phe val arg phe gly lys asn thr OPA ser pro phe
61/21                               91/31
TAT TGG CAA TGC TGG AAA TGG ACA TTC CAA TAT TGC GCG AAT TAA CCG AAC ACG GTG AGG
tyr trp gln cys trp lys trp thr phe gln tyr cys ala asn OCH pro asn thr val arg
121/41                               151/51
GGG GGG CAA GCG TTT GTA CCG GGG CCA GCA AGC GCC GCC GAC CGG TTG ACC GAA GCC AGC
gly gly gln ala phe val pro gly pro ala ser ala ala asp arg leu thr glu ala ser
181/61                               211/71
ATG TTG TTG TGT CAG CGC GGG CTT GGT CTC GAT GTC CCG GCC TTG GCT GGA CCC GCT TCT
met leu leu cys gln arg gly leu gly leu asp val pro ala leu ala gly pro ala ser
241/81                               271/91
TCA AAA CAG GTT GAA CTT AAC GAC TCA AGA ACG GAA ACG CTT GAA CCG CGA CGT CGC TCC
ser lys gln val glu leu asn asp ser arg thr glu thr leu glu pro arg arg arg ser
301/101                               331/111
GGA CAC CAA TTT GAC TCG GCT CTT TGG CAA TTG AAG GTG AGC TGC GAG CAG CCG GGT GAC
gly his gln phe asp ser ala leu trp gln leu lys val ser cys glu gln pro gly asp
361/121                               391/131
CGC ATC GTT GGC CTT GCC ATC AAT CGC CGG CTC GCG GAC GTA GAT AAT CAG CTC ACC GTT
arg ile val gly leu ala ile asn arg arg leu ala asp val asp asn gln leu thr val
421/141                               451/151
GGG ACC GAC CTC GAC CAG GGG TCC TTT GTG ACT GCC GGG CTT GAC GCG GAC GAC CAC AGA
gly thr asp leu asp gln gly ser phe val thr ala gly leu asp ala asp asp his arg
481/161                               511/171
GTC GGT CAT CGC CTA AGG CTA CCG TTC TGA CCT GGG GCT GCG TGG GCG CCG ACG ACG TGA
val gly his arg leu arg leu pro phe OPA pro gly ala ala trp ala pro thr thr OPA
541/181                               571/191
GGC ACG TCA TGT CTC AGC GGC CCA CCG CCA CCT CGG TCG CCG GCA GTA TGT CAG CAT GTG
gly thr ser cys leu ser gly pro pro pro pro arg ser pro ala val cys gln his val
601/201                               631/211
CAG ATG ACT CCA CGC AGC CTT GTT CGC ATC GTT GGT GTC GTG GTT GCG ACG ACC TTG GCG
gln met thr pro arg ser leu val arg ile val gly val val val ala thr thr leu ala
661/221                               691/231
CTG GTG AGC GCA CCC GCC GGC GGT CGT GCC GCG CAT GCG GAT C
leu val ser ala pro ala gly gly arg ala ala his ala asp

```

SEQ ID N° 13A

FIGURE 13A

FEUILLE DE REMPLACEMENT (REGLE 26)

38/185

32/11
 GGA TTT CGT TGC CCG ATG GAT TGT TTG TAC GGT TTG GGA AAA ACA CTT GAA GTC CTT TTT
 gly phe arg cys pro met asp cys leu tyr gly leu gly lys thr leu glu val leu phe
 62/21
 ATT GGC AAT GCT GGA AAT GGA CAT TCC AAT ATT GCG CGA ATT AAC CGA ACA CGG TGA GGG
 ile gly asn ala gly asn gly his ser asn ile ala arg ile asn arg thr arg OPA gly
 122/41
 GGG GGC AAG CGT TTG TAC CGG GGC CAG CAA GCG CCG CCG ACC GGT TGA CCG AAG CCA GCA
 gly gly lys arg leu tyr arg gly gln gln ala pro pro thr gly OPA pro lys pro ala
 182/61
 TGT TGT TGT GTC AGC GCG GGC TTG GTC TCG ATG TCC CGG CCT TGG CTG GAC CCG CTT CTT
 cys cys cys val ser ala gly leu val ser met ser arg pro trp leu asp pro leu leu
 242/81
 CAA AAC AGG TTG AAC TTA ACG ACT CAA GAA CGG AAA CGC TTG AAC CGC GAC GTC GCT CCG
 gln asn arg leu asn leu thr thr gln glu arg lys arg leu asn arg asp val ala pro
 302/101
 GAC ACC AAT TTG ACT CGG CTC TTT GGC AAT TGA AGG TGA GCT GCG AGC AGC CCG GTG ACC
 asp thr asn leu thr arg leu phe gly asn OPA arg OPA ala ala ser ser arg val thr
 362/121
 GCA TCG TTG GCC TTG CCA TCA ATC GCC GGC TCG CGG ACG TAG ATA ATC AGC TCA CCG TTG
 ala ser leu ala leu pro ser ile ala gly ser arg thr AMB ile ile ser ser pro leu
 422/141
 GGA CCG ACC TCG ACC AGG GGT CCT TTG TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG
 gly pro thr ser thr arg gly pro leu OPA leu pro gly leu thr arg thr thr thr glu
 482/161
 TCG GTC ATC GCC TAA GGC TAC CGT TCT GAC CTG GGG CTG CGT GGG CGC CGA CGA CGT GAG
 ser val ile ala OCH gly tyr arg ser asp leu gly leu arg gly arg arg arg arg glu
 542/181
 GCA CGT CAT GTC TCA GCG GCC CAC CGC CAC CTC GGT CGC CGG CAG TAT GTC AGC ATG TGC
 ala arg his val ser ala ala his arg his leu gly arg arg gln tyr val ser met cys
 602/201
 AGA TGA CTC CAC GCA GCC TTG TTC GCA TCG TTG GTG TCG TGG TTG CGA CGA CCT TGG CGC
 arg OPA leu his ala ala leu phe ala ser leu val ser trp leu arg arg pro trp arg
 662/221
 TGG TGA GCG CAC CCG CCG GCG GTC GTG CCG CGC ATG CGG ATC
 trp OPA ala his pro pro ala val val pro arg met arg ile

SEQ ID N° 13B

FIGURE 13B

39/185

33/11

GAT TTC GTT GCC CGA TGG ATT GTT TGT ACG GTT TGG GAA AAA CAC TTG AAG TCC TTT TTA
 asp phe val ala arg trp ile val cys thr val trp glu lys his leu lys ser phe leu
 63/21 93/31

TTG GCA ATG CTG GAA ATG GAC ATT CCA ATA TTG CGC GAA TTA ACC GAA CAC GGT GAG GGG
 leu ala met leu glu met asp ile pro ile leu arg glu leu thr glu his gly glu gly
 123/41 153/51

GGG GCA AGC GTT TGT ACC GGG GCC AGC AAG CGC CGC CGA CCG GTT GAC CGA AGC CAG CAT
 gly ala ser val cys thr gly ala ser lys arg arg arg pro val asp arg ser gln his
 183/61 213/71

GTT GTT GTG TCA GCG CGG GCT TGG TCT CGA TGT CCC GGC CTT GGC TGG ACC CGC TTC TTC
 val val val ser ala arg ala trp ser arg cys pro gly leu gly trp thr arg phe phe
 243/81 273/91

AAA ACA GGT TGA ACT TAA CGA CTC AAG AAC GGA AAC GCT TGA ACC GCG ACG TCG CTC CGG
 lys thr gly OPA thr OCH arg leu lys asn gly asn ala OPA thr ala thr ser leu arg
 303/101 333/111

ACA CCA ATT TGA CTC GGC TCT TTG GCA ATT GAA GGT GAG CTG CGA GCA GCC GGG TGA CCG
 thr pro ile OPA leu gly ser leu ala ile glu gly glu leu arg ala ala gly OPA pro
 363/121 393/131

CAT CGT TGG CCT TGC CAT CAA TCG CCG GCT CGC GGA CGT AGA TAA TCA GCT CAC CGT TGG
 his arg trp pro cys his gln ser pro ala arg gly arg arg OCH ser ala his arg trp
 423/141 453/151

GAC CGA CCT CGA CCA GGG GTC CTT TGT GAC TGC CGG GCT TGA CGC GGA CGA CCA CAG AGT
 asp arg pro arg pro gly val leu cys asp cys arg ala OPA arg gly arg pro gln ser
 483/161 513/171

CGG TCA TCG CCT AAG GCT ACC GTT CTG ACC TGG GGC TGC GTG GGC GCC GAC GAC GTG AGG
 arg ser ser pro lys ala thr val leu thr trp gly cys val gly ala asp asp val arg
 543/181 573/191

CAC GTC ATG TCT CAG CGG CCC ACC GCC ACC TCG GTC GCC GGC AGT ATG TCA GCA TGT GCA
 his val met ser gln arg pro thr ala thr ser val ala gly ser met ser ala cys ala
 603/201 633/211

GAT GAC TCC ACG CAG CCT TGT TCG CAT CGT TGG TGT CGT GGT TGC GAC GAC CTT GGC GCT
 asp asp ser thr gln pro cys ser his arg trp cys arg gly cys asp asp leu gly ala
 663/221 693/231

GGT GAG CGC ACC CGC CGG CGG TCG TGC CGC GCA TGC GGA TC
 gly glu arg thr arg arg arg ser cys arg ala cys gly

SEQ ID N° 13C

FIGURE 13C

40/185

partie de la séquence nucléotidique de seq13A

```

1/1                               31/11
GGG TCC TTT GTG ACT GCC GGG CTT GAC GCG GAC GAC CAC AGA GTC GGT CAT CGC CTA AGG
gly ser phe val thr ala gly leu asp ala asp asp his arg val gly his arg leu arg
61/21                               91/31
CTA CCG TTC TGA CCT GGG GCT GCG TGG GCG CCG ACG ACG TGA GGC ACG TCA TGT CTC AGC
leu pro phe OPA pro gly ala ala trp ala pro thr thr OPA gly thr ser cys leu ser
121/41                               151/51
GGC CCA CCG CCA CCT CGG TCG CCG GCA GTA TGT CAG CAT GTG CAG ATG ACT CCA CGC AGC
gly pro pro pro pro arg ser pro ala val cys gln his val gln met thr pro arg ser
181/61                               211/71
CTT GTT CGC ATC GTT GGT GTC GTG GTT GCG ACG ACC TTG GCG CTG GTG AGC GCA CCC GCC
leu val arg ile val gly val val val ala thr thr leu ala leu val ser ala pro ala
241/81
GGC GGT CGT GCC GCG CAT GCG GAT C
gly gly arg ala ala his ala asp

```

SEQ ID N° 13A'

FIGURE 13A'

```

1/1                               31/11
GGT CCT TTG TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG TCG GTC ATC GCC TAA GGC
gly pro leu OPA leu pro gly leu thr arg thr thr thr glu ser val ile ala OCH gly
61/21                               91/31
TAC CGT TCT GAC CTG GGG CTG CGT GGG CGC CGA CGA CGT GAG GCA CGT CAT GTC TCA CGC
tyr arg ser asp leu gly leu arg gly arg arg arg arg glu ala arg his val ser ala
121/41                               151/51
GCC CAC CGC CAC CTC GGT CGC CGG CAG TAT GTC AGC ATG TGC AGA TGA CTC CAC GCA GCC
ala his arg his leu gly arg arg gln tyr val ser met cys arg OPA leu his ala ala
181/61                               211/71
TTG TTC GCA TCG TTG GTG TCG TGG TTG CGA CGA CCT TGG CGC TGG TGA GCG CAC CCG CCG
leu phe ala ser leu val ser trp leu arg arg pro trp arg trp OPA ala his pro pro
241/81
GCG GTC GTG CCG CGC ATG CGG ATC
ala val val pro arg met arg ile

```

SEQ ID N° 13B'

FIGURE 13B'

```

1/1                               31/11
GTC CTT TGT GAC TGC CGG GCT TGA CGC GGA CGA CCA CAG AGT CGG TCA TCG CCT AAG GCT
val leu cys asp cys arg ala OPA arg gly arg pro gln ser arg ser ser pro lys ala
61/21                               91/31
ACC GTT CTG ACC TGG GGC TGC GTG GGC GCC GAC GAC GTG AGG CAC GTC ATG TCT CAG CGG
thr val leu thr trp gly cys val gly ala asp asp val arg his val met ser gln arg
121/41                               151/51
CCC ACC GCC ACC TCG GTC GCC GGC AGT ATG TCA GCA TGT GCA GAT GAC TCC ACG CAG CCT
pro thr ala thr ser val ala gly ser met ser ala cys ala asp asp ser thr gln pro
181/61                               211/71
TGT TCG CAT CGT TGG TGT CGT GGT TGC GAC GAC CTT GGC GCT GGT GAG CGC ACC CGC CGG
cys ser his arg trp cys arg gly cys asp asp leu gly ala gly glu arg thr arg arg
241/81
CGG TCG TGC CGC GCA TGC GGA TC
arg ser cys arg ala cys gly

```

SEQ ID N° 13C'

FIGURE 13C'

FEUILLE DE REMPLACEMENT (REGLE 26)

41/185

séquence Rv1984c prédite par Cole et al. (Nature 393:537-544) et contenant seq13A'

```

1/1                               31/11
atg act cca cgc agc ctt gtt cgc atc gtt ggt gtc gtg gtt gcg acg acc ttg gcg ctg
Met thr pro arg ser leu val arg ile val gly val val val ala thr thr leu ala leu
61/21                               91/31
gtg agc gca ccc gcc ggc ggt cgt gcc gcg cat gcg gat ccg tgt tcg gac atc gcg gtc
val ser ala pro ala gly gly arg ala ala his ala asp pro cys ser asp ile ala val
121/41                              151/51
gtt ttc gct cgc ggc acg cat cag gct tct ggt ctt ggc gac gtc ggt gag gcg ttc gtc
val phe ala arg gly thr his gln ala ser gly leu gly asp val gly glu ala phe val
181/61                              211/71
gac tcg ctt acc tcg caa gtt ggc ggg cgg tcg att ggg gtc tac gcg gtg aac tac cca
asp ser leu thr ser gln val gly gly arg ser ile gly val tyr ala val asn tyr pro
241/81                              271/91
gca agc gac gac tac cgc gcg agc gcg tca aac ggt tcc gat gat gcg agc gcc cac atc
ala ser asp asp tyr arg ala ser ala ser asn gly ser asp asp ala ser ala his ile
301/101                             331/111
cag cgc acc gtc gcc agc tgc ccg aac acc agg att gtg ctt ggt ggc tat tcg cag ggt
gln arg thr val ala ser cys pro asn thr arg ile val leu gly gly tyr ser gln gly
361/121                             391/131
gcg acg gtc atc gat ttg tcc acc tcg gcg atg ccg ccc gcg gtg gca gat cat gtc gcc
ala thr val ile asp leu ser thr ser ala met pro pro ala val ala asp his val ala
421/141                             451/151
gct gtc gcc ctt ttc ggc gag cca tcc agt ggt ttc tcc agc atg ttg tgg ggc ggc ggg
ala val ala leu phe gly glu pro ser ser gly phe ser ser met leu trp gly gly gly
481/161                             511/171
tcg ttg ccg aca atc ggt ccg ctg tat agc tct aag acc ata aac ttg tgt gct ccc gac
ser leu pro thr ile gly pro leu tyr ser ser lys thr ile asn leu cys ala pro asp
541/181                             571/191
gat cca ata tgc acc gga ggc ggc aat att atg gcg cat gtt tcg tat gtt cag tcg ggg
asp pro ile cys thr gly gly gly asn ile met ala his val ser tyr val gln ser gly
601/201                             631/211
atg aca agc cag gcg gcg aca ttc gcg gcg aac agg ctc gat cac gcc gga tga
met thr ser gln ala ala thr phe ala ala asn arg leu asp his ala gly OPA

```

SEQ ID N° 13D

FIGURE 13D

42/185

Seq13F: ORF prédite par Cole et al. (Nature 393:537-544) et contenant Rv1984c

```

1/1                                31/11
tga ggc acg tca tgt ctc agc ggc cca ccg cca cct cgg tcg ccg gca gta tgt cag cat
OPA gly thr ser cys leu ser gly pro pro pro pro arg ser pro ala val cys gln his
61/21                                91/31
gtg cag atg act cca cgc agc ctt gtt cgc atc gtt ggt gtc gtg gtt gcg acg acc ttg
val gln met thr pro arg ser leu val arg ile val gly val val val ala thr thr leu
121/41                               151/51
gcg ctg gtg agc gca ccc gcc ggc ggt cgt gcc gcg cat gcg gat ccg tgt tcg gac atc
ala leu val ser ala pro ala gly gly arg ala ala his ala asp pro cys ser asp ile
181/61                               211/71
gcg gtc gtt ttc gct cgc ggc acg cat cag gct tct ggt ctt ggc gac gtc ggt gag gcg
ala val val phe ala arg gly thr his gln ala ser gly leu gly asp val gly glu ala
241/81                               271/91
ttc gtc gac tcg ctt acc tcg caa gtt ggc ggg cgg tcg att ggg gtc tac gcg gtg aac
phe val asp ser leu thr ser gln val gly gly arg ser ile gly val tyr ala val asn
301/101                              331/111
tac cca gca agc gac gac tac cgc gcg agc gcg tca aac ggt tcc gat gat gcg agc gcc
tyr pro ala ser asp asp tyr arg ala ser ala ser asn gly ser asp asp ala ser ala
361/121                              391/131
cac atc cag cgc acc gtc gcc agc tgc ccg aac acc agg att gtg ctt ggt ggc tat tcg
his ile gln arg thr val ala ser cys pro asn thr arg ile val leu gly gly tyr ser
421/141                              451/151
cag ggt gcg acg gtc atc gat ttg tcc acc tcg gcg atg ccg ccc gcg gtg gca gat cat
gln gly ala thr val ile asp leu ser thr ser ala met pro pro ala val ala asp his
481/161                              511/171
gtc gcc gct gtc gcc ctt ttc ggc gag cca tcc agt ggt ttc tcc agc atg ttg tgg ggc
val ala ala val ala leu phe gly glu pro ser ser gly phe ser ser met leu trp gly
541/181                              571/191
ggc ggg tcg ttg ccg aca atc ggt ccg ctg tat agc tct aag acc ata aac ttg tgt gct
gly gly ser leu pro thr ile gly pro leu tyr ser ser lys thr ile asn leu cys ala
601/201                              631/211
ccc gac gat cca ata tgc acc gga ggc ggc aat att atg gcg cat gtt tcg tat gtt cag
pro asp asp pro ile cys thr gly gly gly asn ile met ala his val ser tyr val gln
661/221                              691/231
tcg ggg atg aca agc cag gcg gcg aca ttc gcg gcg aac agg ctc gat cac gcc gga tga
ser gly met thr ser gln ala ala thr phe ala ala asn arg leu asp his ala gly OPA

```

SEQ ID N° 13F

FIGURE 13F

43/185

31/11
CCA CCG GGG CTG GAG GGG CGA ATG TGC GCC GAA CGC CGT CGG CCA ACT TGG CCG CTG AGG
pro pro gly leu glu gly arg met cys ala glu arg arg arg pro thr trp pro leu arg
61/21
GCG GCT GAT CCC CTG GCC CGA GAC GGG GCA AGC CAA TAG CGG CTC CAT CGG GCT TTG CTG
ala ala asp pro leu ala arg asp gly ala ser gln AMB arg leu his arg ala leu leu
121/41
GTA GCG GTT CGG CGG GAA CCG AGC GCC GAC GTT GTC GGT GCC CGG TGA TAT ATT GGG TCA
val ala val arg arg glu pro ser ala asp val val gly ala arg OPA tyr ile gly ser
181/61
GAC GGG TAT GGC GGC GAC TGA GGT GAT CTG CGA CAC GCC GCC GCG GTG CTC GAG CCA GGC
asp gly tyr gly gly asp OPA gly asp leu arg his ala ala val leu glu pro gly
241/81
TTA CGA CCA GGG AAT TTC GAA AAT GTT ATT CAG AAC ATC TTG TAT CTC TTC CTC CGT GCC
leu arg pro gly asn phe glu asn val ile gln asn ile leu tyr leu phe leu arg ala
301/101
ACC CCC TAG GTG TAG TGT TTT CGA GTA CCG GCA GAT CCC AGT TCA CCA GTC TCA CCA GAT
thr pro AMB val AMB cys phe arg val pro ala asp pro ser ser pro val ser pro asp

C

SEQ ID N° 14A

FIGURE 14A

32/11
CAC CGG GGC TGG AGG GGC GAA TGT GCG CCG AAC GCC GTC GGC CAA CTT GGC CGC TGA GGG
his arg gly trp arg gly glu cys ala pro asn ala val gly gln leu gly arg OPA gly
62/21
CGG CTG ATC CCC TGG CCC GAG ACG GGG CAA GCC AAT AGC GGC TCC ATC GGG CTT TGC TGG
arg leu ile pro trp pro glu thr gly gln ala asn ser gly ser ile gly leu cys trp
122/41
TAG CGG TTC GGC GGG AAC CGA GCG CCG ACG TTG TCG GTG CCC GGT GAT ATA TTG GGT CAG
AMB arg phe gly gly asn arg ala pro thr leu ser val pro gly asp ile leu gly gln
182/61
ACG GGT ATG GCG GCG ACT GAG GTG ATC TGC GAC ACG CCG CCG CGG TGC TCG AGC CAG GCT
thr gly met ala ala thr glu val ile cys asp thr pro pro arg cys ser ser gln ala
242/81
TAC GAC CAG GGA ATT TCG AAA ATG TTA TTC AGA ACA TCT TGT ATC TCT TCC TCC GTG CCA
tyr asp gln gly ile ser lys met leu phe arg thr ser cys ile ser ser ser val pro
302/101
CCC CCT AGG TGT AGT GTT TTC GAG TAC CCG CAG ATC CCA GTT CAC CAG TCT CAC CAG ATC
pro pro arg cys ser val phe glu tyr arg gln ile pro val his gln ser his gln ile

SEQ ID N° 14B

FIGURE 14B

44/185

33/11
 ACC GGG GCT GGA GGG GCG AAT GTG CGC CGA ACG CCG TCG GCC AAC TTG GCC GCT GAG GGC
 thr gly ala gly gly ala asn val arg arg thr pro ser ala asn leu ala ala glu gly
 63/21
 GGC TGA TCC CCT GGC CCG AGA CGG GGC AAG CCA ATA GCG GCT CCA TCG GGC TTT GCT GGT
 gly OPA ser pro gly pro arg arg gly lys pro ile ala ala pro ser gly phe ala gly
 123/41
 AGC GGT TCG GCG GGA ACC GAG CGC CGA CGT TGT CGG TGC CCG GTG ATA TAT TGG GTC AGA
 ser gly ser ala gly thr glu arg arg arg cys arg cys pro val ile tyr trp val arg
 183/61
 CGG GTA TGG CGG CGA CTG AGG TGA TCT GCG ACA CGC CGC CGC GGT GCT CGA GCC AGG CTT
 arg val trp arg arg leu arg OPA ser ala thr arg arg arg gly ala arg ala arg leu
 243/81
 ACG ACC AGG GAA TTT CGA AAA TGT TAT TCA GAA CAT CTT GTA TCT CTT CCT CCG TGC CAC
 thr thr arg glu phe arg lys cys tyr ser glu his leu val ser leu pro pro cys his
 303/101
 CCC CTA GGT GTA GTG TTT TCG AGT ACC GGC AGA TCC CAG TTC ACC AGT CTC ACC AGA TC
 pro leu gly val val phe ser ser thr gly arg ser gln phe thr ser leu thr arg

SEQ ID N° 14C

FIGURE 14C

partie de la séquence nucléotidique de seq14A

1/1 31/11
 TTT TCG AGT ACC GGC AGA TCC CAG GTT CAC CAG GTC TCA CCA GAT C
 phe ser ser thr gly arg ser gln val his gln val ser pro asp

SEQ ID N° 14A'

FIGURE 14A'

1/1 31/11
 TGT TTT CGA GTA CCG GCA GAT CCC AGG TTC ACC AGG TCT CAC CAG ATC
 cys phe arg val pro ala asp pro arg phe thr arg ser his gln ile

SEQ ID N° 14C

FIGURE 14C

1/1 31/11
 GTT TTC GAG TAC CGG CAG ATC CCA GGT TCA CCA GGT CTC ACC AGA TC
 val phe glu tyr arg gln ile pro gly ser pro gly leu thr arg

SEQ ID N° 14C'

FIGURE 14C'

FEUILLE DE REMPLACEMENT (REGLE 26)

45/185

ORF prédite d'après la séquence publiée par Cole et al. (Nature 393:537-544) et contenant seq14A'

```

1/1                                31/11
TAG CGG TTC GGC GGG AAG CTA GCG GCG ACG TTG TCG GTG GCC GGT GAT ATA TTG GGT CAG
AMB arg phe gly gly lys leu ala ala thr leu ser val ala gly asp ile leu gly gln
61/21                                91/31
ACG GGT ATG GCG GCG GCT GAG GTG ATC TGC GAC ACG CCG CCG CGG TGC TCG AGC CAG GCT
thr gly met ala ala ala glu val ile cys asp thr pro pro arg cys ser ser gln ala
121/41                                151/51
TAC GAC CAG GGA ATT TCG AAA ATG TTA TTC AGA ACA TCT TGT ATC TCT TCT CCG TGC CAC
tyr asp gln gly ile ser lys met leu phe arg thr ser cys ile ser ser pro cys his
181/61                                211/71
CCC CTA GGT GTA GTG TTT TCG AGT ACC GGC AGA TCC CAG GTT CAC CAG GTC TCA CCA gat
pro leu gly val val phe ser ser thr gly arg ser gln val his gln val ser pro asp
241/81                                271/91
cca cgg ggc gcg atg aac ttc ccg gca tcg gca tcg cca ggt cga cgg acg tgg tcg cgc
pro arg gly ala met asn phe pro ala ser ala ser pro gly arg arg thr trp ser arg
301/101                                331/111
tat gac ggg aat ctg gag cct tgt cgg gcc gct caa cat atc gaa gat gca cta ctt gag
tyr asp gly asn leu glu pro cys arg ala ala gln his ile glu asp ala leu leu glu
361/121                                391/131
tcg ttg cca gat cct gtc aga ttc ccg att tcc gca aag gag cgg tac gcc cat gac cgt
ser leu pro asp pro val arg phe pro ile ser ala lys glu arg tyr ala his asp arg
421/141
gac cgt tta cac taa
asp arg leu his OCH

```

SEQ ID N° 14F

FIGURE 14F

46/185

Séquence Rv3054c prédite par Cole et l. (Nature 393:537-544)
pouvant être en phase avec Seq14A'

1/1	31/11
gtg tca gat acc aag tcc gac atc aaa atc	ttg gcc tta gtg gga agc ctg cgc gcg gcg
val ser asp thr lys ser asp ile lys ile	leu ala leu val gly ser leu arg ala ala
61/21	91/31
tcg ttc aac cgc cag atc gcc gag ctg gct	gcc aag gtc gct ccg gac gcc gtc acc gtc
ser phe asn arg gln ile ala glu leu ala	ala lys val ala pro asp gly val thr val
121/41	151/51
acc atg ttc gag ggg ctg ggg gac ctg ccg	ttc tac aac gaa gac atc gac aca gcg acg
thr met phe glu gly leu gly asp leu pro	phe tyr asn glu asp ile asp thr ala thr
181/61	211/71
gag gtg ccg gcg ccg gtg agc gcg ttg ccg	gag gcc gcg tct gac gcg cac gct gcc ttg
glu val pro ala pro val ser ala leu arg	glu ala ala ser asp ala his ala ala leu
241/81	271/91
gtg gtc acg ccg gaa tac aac gcc agc att	ccg gcc gtg atc aag aac gcg atc gac tgg
val val thr pro glu tyr asn gly ser ile	pro ala val ile lys asn ala ile asp trp
301/101	331/111
ctg tcc agg cca ttc gcc gat gcc gcg ttg	aag gac aag ccg ttg gcc gtg atc gcc gcc
leu ser arg pro phe gly asp gly ala leu	lys asp lys pro leu ala val ile gly gly
361/121	391/131
tcc atg gcc cgc tac gcc ggg gta tgg gcg	cac gac gag act cgc aag tcg ttc agc atc
ser met gly arg tyr gly gly val trp ala	his asp glu thr arg lys ser phe ser ile
421/141	451/151
gct gcc acg cgg gtg gtc gat gcg atc aaa	ctg tcg gtg ccg ttc caa act ctg gcc aag
ala gly thr arg val val asp ala ile lys	leu ser val pro phe gln thr leu gly lys
481/161	511/171
tcg gtc gcg gac gac gcc ggg ctg gcg gcg	aat gtg cgc gac gcc gtc gcc aac ttg gcc
ser val ala asp asp ala gly leu ala ala	asn val arg asp ala val gly asn leu ala
541/181	
gct gag gtc gcc tga	
ala glu val gly OPA	

SEQ ID N° 14R

FIGURE 14R

47/185

ORF prédite par Cole et al. (Nature 393:537-544) et contenant Rv3054c

```

1/1                                31/11
taa cgc gat cgg aat aaa tcg gac cat ggt ccg gtt ggc tcg tgc aag gac gtg gac caa
OCH arg asp arg asn lys ser asp his gly pro val gly ser cys lys asp val asp gln
61/21                                91/31
caa gcg gaa agg aac gta gca gtg tca gat acc aag tcc gac atc aaa atc ttg gcc tta
gln ala glu arg asn val ala val ser asp thr lys ser asp ile lys ile leu ala leu
121/41                                151/51
gtg gga agc ctg cgc gcg gcg tcg ttc aac cgc cag atc gcc gag ctg gct gcc aag gtc
val gly ser leu arg ala ala ser phe asn arg gln ile ala glu leu ala ala lys val
181/61                                211/71
gct ccg gac ggc gtc acc gtc acc atg ttc gag ggg ctg ggg gac ctg ccg ttc tac aac
ala pro asp gly val thr val thr met phe glu gly leu gly asp leu pro phe tyr asn
241/81                                271/91
gaa gac atc gac aca gcg acg gag gtg ccg gcg ccg gtg agc gcg ttg ccg gag gcc gcg
glu asp ile asp thr ala thr glu val pro ala pro val ser ala leu arg glu ala ala
301/101                               331/111
tct gac gcg cac gct gcc ttg gtg gtc acg ccg gaa tac aac ggc agc att ccg gcc gtg
ser asp ala his ala ala leu val val thr pro glu tyr asn gly ser ile pro ala val
361/121                               391/131
atc aag aac gcg atc gac tgg ctg tcc agg cca ttc ggc gat ggc gcg ttg aag gac aag
ile lys asn ala ile asp trp leu ser arg pro phe gly asp gly ala leu lys asp lys
421/141                               451/151
ccg ttg gcc gtg atc ggc ggc tcc atg ggc cgc tac ggc ggg gta tgg gcg cac gac gag
pro leu ala val ile gly gly ser met gly arg tyr gly gly val trp ala his asp glu
481/161                               511/171
act cgc aag tcg ttc agc atc gct ggc acg ccg gtg gtc gat gcg atc aaa ctg tcg gtg
thr arg lys ser phe ser ile ala gly thr arg val val asp ala ile lys leu ser val
541/181                               571/191
ccg ttc caa act ctg ggc aag tcg gtc gcg gac gac gcc ggg ctg gcg gcg aat gtg cgc
pro phe gln thr leu gly lys ser val ala asp asp ala gly leu ala ala asn val arg
601/201                               631/211
gac gcc gtc ggc aac ttg gcc gct gag gtc ggc tga
asp ala val gly asn leu ala ala glu val gly OPA

```

SEQ ID N° 14P

FIGURE 14P

48/185

fragment d'après la séquence publiée par Cole et al. (Nature 393:537-544) et contenant seq 14F' et seq 14P'

1/1
 taa cgc gat cgg aat aaa tcg gac cat ggt ccg gtt ggc tcg tgc aag gac gtg gac caa
 OCH arg asp arg asn lys ser asp his gly pro val gly ser cys lys asp val asp gln
 asn ala ile gly ile asn arg thr met val arg leu ala arg ala arg thr trp thr asn
 thr arg ser glu OCH ile gly pro trp ser gly trp leu val gln gly arg gly pro thr
 61/21
 caa gcg gaa agg aac gta gca gtg tca gat acc aag tcc gac atc aaa atc ttg gcc tta
 gln ala glu arg asn val ala val ser asp thr lys ser asp ile lys ile leu ala leu
 lys arg lys gly thr AMB gln cys gln ile pro ser pro thr ser lys ser trp pro AMB
 ser gly lys glu arg ser ser val arg tyr gln val arg his gln asn leu gly leu ser
 121/41
 gtg gga agc ctg cgc gcg gcg tcg ttc aac cgc cag atc gcc gag ctg gct gcc aag gtc
 val gly ser leu arg ala ala ser phe asn arg gln ile ala glu leu ala ala lys val
 trp glu ala cys ala arg arg arg ser thr ala arg ser pro ser trp leu pro arg ser
 gly lys pro ala arg gly val val gln pro pro asp arg arg ala gly cys gln gly arg
 181/61
 gct ccg gac ggc gtc acc gtc acc atg ttc gag ggg ctg ggg gac ctg ccg ttc tac aac
 ala pro asp gly val thr val thr met phe glu gly leu gly asp leu pro phe tyr asn
 leu arg thr ala ser pro ser pro cys ser arg gly trp gly thr cys arg ser thr thr
 ser gly arg arg his arg his his val arg gly ala gly pro ala val leu gln arg
 241/81
 gaa gac atc gac aca gcg acg gag gtg ccg gcg ccg gtg agc gcg ttg ccg gag gcc gcg
 glu asp ile asp thr ala thr glu val val thr pro glu tyr asn gly ser ile pro ala val
 lys thr ser thr gln arg arg cys arg arg OPA ala arg cys gly arg pro arg
 arg his arg his ser asp gly gly ala gly ala gly glu arg val ala gly gly arg val
 301/101
 tct gac gcg cac gct gcc ttg gtg gtc acg ccg gaa tac aac ggc agc att ccg gcc gtg
 ser asp ala his ala ala leu val val thr pro glu tyr asn gly ser ile pro ala val
 leu thr arg thr leu pro trp trp ser arg arg asn thr thr ala ala phe arg pro OPA
 OPA arg ala arg cys leu gly gly his ala gly ile gln arg gln his ser gly arg asp
 361/121
 atc aag aac gcg atc gac tgg ctg tcc agg cca ttc ggc gat ggc gcg ttg aag gac aag
 ile lys asn ala ile asp trp leu ser arg pro phe gly asp gly ala leu lys asp lys
 ser arg thr arg ser thr gly cys pro gly his ser ala met ala arg OPA arg thr ser
 gln glu arg asp arg leu ala val gln ala ile arg arg trp arg val glu gly gln ala
 421/141
 ccg ttg gcc gtg atc ggc ggc tcc atg ggc cgc tac ggc ggg gta tgg gcg cac gac gag
 pro leu ala val ile gly gly ser met gly arg tyr gly gly val trp ala his asp glu
 arg trp pro OPA ser ala ala pro trp ala ala thr ala gly tyr gly arg thr thr arg
 val gly arg asp arg arg leu his gly pro leu arg arg gly met gly ala arg arg asp
 481/161
 act cgc aag tcg ttc agc atc gct ggc acg cgg gtg gtc gat gcg atc aaa ctg tcg gtg
 thr arg lys ser phe ser ile ala gly thr arg val val asp ala ile lys leu ser val
 leu ala ser arg ser ala ser leu ala arg gly trp ser met arg ser asn cys arg cys
 ser gln val val gln his arg trp his ala gly gly arg cys asp gln thr val gly ala
 541/181
 ccg ttc caa act ctg ggc aag tcg gtc gcg gac gac gcc ggg ctg gcg gcg aat gtg cgc
 pro phe gln thr leu gly lys ser val ala asp asp ala gly leu ala ala asn val arg
 arg ser lys leu trp ala ser arg ser arg thr thr pro gly trp arg arg met cys ala
 val pro asn ser gly gln val gly arg gly arg arg ala gly gly glu cys ala arg
 601/201
 gac gcc gtc ggc aac ttg gcc gct gag gtc ggc tga tcc ctg ggc cga ggc ggg tca gcc
 asp ala val gly asn leu ala ala glu val gly OPA ser leu gly arg gly gly ser ala
 thr pro ser ala thr trp pro leu arg ser ala asp pro trp ala glu ala gly gln pro
 arg arg arg gln leu gly arg OPA gly arg leu ile pro gly pro arg arg val ser gln
 661/221
 aat agc ggc tcc atc ggc ttt gct ggt agc ggt tcg gcg gga agc tag ccg cga cgt tgt
 asn ser gly ser ile gly phe ala gly ser gly ser ala gly ser AMB arg arg arg cys
 ile ala ala pro ser ala leu leu val ala val arg arg glu ala ser gly asp val val
 AMB arg leu his arg leu cys trp AMB arg phe gly gly lys leu ala ala thr leu ser

SEQ ID N° 14Q

FEUILLE DE REMPLACEMENT (REGLE 26)

49/185

721/241 751/251
 cgg tgg ccg gtg ata tat tgg gtc aga cgg gta tgg cgg cgg ctg agg tga tct gcg aca
 arg trp pro val ile tyr trp val arg arg val trp arg arg leu arg OPA ser ala thr
 gly gly arg OPA tyr ile gly ser asp gly tyr gly gly OPA gly asp leu arg his
 val ala gly asp ile leu gly gln thr gly met ala ala ala glu val ile cys asp thr
 781/261 811/271
 cgc cgc cgc ggt gct cga gcc agg ctt acg acc agg gaa ttt cga aaa tgt tat tca gaa
 arg arg arg gly ala arg ala arg leu thr thr arg glu phe arg lys cys tyr ser glu
 ala ala ala val leu glu pro gly leu arg pro gly asn phe glu asn val ile gln asn
 pro pro arg cys ser ser gln ala tyr asp gln gly ile ser lys met leu phe arg thr
 841/281 871/291
 cat ctt gta tct ctt ctc cgt gcc acc ccc tag gtg tag tgt ttt cga gta ccg gca gat
 his leu val ser leu leu arg ala thr pro AMB val AMB cys phe arg val pro ala asp
 ile leu tyr leu phe ser val pro pro pro arg cys ser val phe glu tyr arg gln ile
 ser cys ile ser ser pro cys his pro leu gly val val phe ser ser thr gly arg ser
 901/301 931/311
 ccc agg ttc acc agg tct cac cag atc cac ggg gcg cga tga act tcc cgg cat cgg cat
 pro arg phe thr arg ser his gln ile his gly ala arg OPA thr ser arg his arg his
 pro gly ser pro gly leu thr arg ser thr gly arg asp glu leu pro gly ile gly ile
 gln val his gln val ser pro asp pro arg gly ala met asn phe pro ala ser ala ser
 961/321 991/331
 cgc cag gtc gac gga cgt ggt cgc gct atg acg gga atc tgg agc ctt gtc ggg ccg ctc
 arg gln val asp gly arg gly arg ala met thr gly ile trp ser leu val gly pro leu
 ala arg ser thr asp val val ala leu OPA arg glu ser gly ala leu ser gly arg ser
 pro gly arg arg thr trp ser arg tyr asp gly asn leu glu pro cys arg ala ala gln
 1021/341 1051/351
 aac ata tcg aag atg cac tac ttg agt cgt tgc cag atc ctg tca gat tcc cga ttt ccg
 asn ile ser lys met his tyr leu ser arg cys gln ile leu ser asp ser arg phe pro
 thr tyr arg arg cys thr thr OPA val val ala arg ser cys gln ile pro asp phe arg
 his ile glu asp ala leu leu glu ser leu pro asp pro val arg phe pro ile ser ala
 1081/361 1111/371
 caa agg agc ggt acg ccc atg acc gtg acc gtt tac act aa
 gln arg ser gly thr pro met thr val thr val tyr thr
 lys gly ala val arg pro OPA pro OPA pro phe thr leu
 lys glu arg tyr ala his asp arg asp arg leu his OCH

SEQ ID N° 14Q(suite)

FIGURE 14Q(suite)

1/1 31/11
 CAA GCC CGG CCG CGA CTG TTT GCC GTT TTG GGG CTC CTA CCA GAA CAC CAC CTG GCG GCC
 gln ala arg pro arg leu phe ala val leu gly leu leu pro glu his his leu ala ala
 61/21 91/31
 GCG CAC CAT GGT GTG CAC CAG TTG CGA TCG GTT CCT CCC GCG CGC GGG CGG CGA CGA CGT
 ala his his gly val his gln leu arg ser val pro pro ala arg gly arg arg arg
 121/41 151/51
 CGA TGC CCG CGC CCC GGC GGC GCA GCT GCG TAG CTC GAC CCG GTC GAC GAC GAC GGG GTC
 arg cys pro arg pro gly gly ala ala ala AMB leu asp pro val asp asp asp gly val
 181/61 211/71
 GGC GGA CCA GTC GGC GAT GTC GAG GCG ATG GCA ATA CAG CGC CTT GGT GCG CGG CCA CAC
 gly gly pro val gly asp val glu ala met ala ile gln arg leu gly ala arg pro his
 241/81 271/91
 GTC TGA GGT GGC GAA GAC CAG TCC CGC GCC CAC CGG CAG CCG GAT CCG GAT ACG CGG TAC
 val OPA gly gly glu asp gln ser arg ala his arg gln pro asp pro asp thr arg tyr

SEQ ID N° 15A

FEUILLE DE REMPLACEMENT (REGLE 26)

50/185

32/11
 AAG CCC GGC CGC GAC TGT TTG CCG TTT TGG GGC TCC TAC CAG AAC ACC ACC TGG CGG CCG
 lys pro gly arg asp cys leu pro phe trp gly ser tyr gln asn thr thr trp arg pro
 62/21
 CGC ACC ATG GTG TGC ACC AGT TGC GAT CGG TTC CTC CCG CGC GCG GGC GGC GAC GAC GTC
 arg thr met val cys thr ser cys asp arg phe leu pro arg ala gly gly asp asp val
 122/41
 GAT GCC CGC GCC CCG GCG GCG CAG CTG CGT AGC TCG ACC CGG TCG ACG ACG ACG GGG TCG
 asp ala arg ala pro ala ala gln leu arg ser ser thr arg ser thr thr thr gly ser
 182/61
 GCG GAC CAG TCG GCG ATG TCG AGG CGA TGG CAA TAC AGC GCC TTG GTG CGC GGC CAC ACG
 ala asp gln ser ala met ser arg arg trp gln tyr ser ala leu val arg gly his thr
 242/81
 TCT GAG GTG GCG AAG ACC AGT CCC GCG CCC ACC GGC AGC CGG ATC CGG ATA CGC GGT AC
 ser glu val ala lys thr ser pro ala pro thr gly ser arg ile arg ile arg gly

SEQ ID N° 15B

FIGURE 15B

33/11
 AGC CCG GCC GCG ACT GTT TGC CGT TTT GGG GCT CCT ACC AGA ACA CCA CCT GGC GGC CGC
 ser pro ala ala thr val cys arg phe gly ala pro thr arg thr pro pro gly gly arg
 63/21
 GCA CCA TGG TGT GCA CCA GTT GCG ATC GGT TCC TCC CGC GCG CGG GCG GCG ACG ACG TCG
 ala pro trp cys ala pro val ala ile gly ser ser arg ala arg ala ala thr thr ser
 123/41
 ATG CCC GCG CCC CGG CGG CGC AGC TGC GTA GCT CGA CCC GGT CGA CGA CGA CGG GGT CGG
 met pro ala pro arg arg arg ser cys val ala arg pro gly arg arg arg arg gly arg
 183/61
 CGG ACC AGT CGG CGA TGT CGA GGC GAT GGC AAT ACA GCG CCT TGG TGC GCG GCC ACA CGT
 arg thr ser arg arg cys arg gly asp gly asn thr ala pro trp cys ala ala thr arg
 243/81
 CTG AGG TGG CGA AGA CCA GTC CCG CGC CCA CCG GCA GCC GGA TCC GGA TAC GCG GTA C
 leu arg trp arg arg pro val pro arg pro pro ala ala gly ser gly tyr ala val

SEQ ID N° 15C

FIGURE 15C

51/185

partie de la séquence nucléotidique de seq15A

1/1	31/11
GGC GGC CGC GCG CCA TGG TGT GCA CCA GTT	GCG ATC GGT TCT CCC GCG CGC GGG CGG CGA
gly gly arg ala pro trp cys ala pro val	ala ile gly ser pro ala arg gly arg arg
61/21	91/31
CGA CGT CGA TGG CCG CGC CCC GGC GGC TGC	AGC TGC GTA GCT CGA CCC GGT CGA CGA CGA
arg arg arg trp pro arg pro gly gly cys	ser cys val ala arg pro gly arg arg arg
121/41	151/51
CGG GGT CGG CGG GCC AGT CGG CGA TGT CGA	GGC GAT GGC AAT ACA GCG CCT TGG TGC GCG
arg gly arg arg ala ser arg arg cys arg	gly asp gly asn thr ala pro trp cys ala
181/61	211/71
GCC ACA CGT CTG AGG TGG CGA AGA CCA GTC	CCG CGC CCA CCG GCA GCC GGA TC
ala thr arg leu arg trp arg arg pro val	pro arg pro pro ala ala gly

SEQ ID N° 15A'

FIGURE 15A'

1/1	31/11
GCG GCC GCG CGC CAT GGT GTG CAC CAG TTG	CGA TCG GTT CTC CCG CGC GCG GGC GGC GAC
ala ala ala arg his gly val his gln leu	arg ser val leu pro arg ala gly gly asp
61/21	91/31
GAC GTC GAT GGC CGC GCC CCG GCG GCT GCA	GCT GCG TAG CTC GAC CCG GTC GAC GAC GAC
asp val asp gly arg ala pro ala ala ala	ala ala AMB leu asp pro val asp asp asp
121/41	151/51
GGG GTC GGC GGG CCA GTC GGC GAT GTC GAG	GCG ATG GCA ATA CAG CGC CTT GGT GCG CGG
gly val gly gly pro val gly asp val glu	ala met ala ile gln arg leu gly ala arg
181/61	211/71
CCA CAC GTC TGA GGT GGC GAA GAC CAG TCC	CGC GCC CAC CGG CAG CCG GAT C
pro his val OPA gly gly glu asp gln ser	arg ala his arg gln pro asp

SEQ ID N° 15B'

FIGURE 15B'

1/1	31/11
TGG CGG CCG CGC GCC ATG GTG TGC ACC AGT	TGC GAT CGG TTC TCC CGC GCG CGG GCG GCG
trp arg pro arg ala met val cys thr ser	cys asp arg phe ser arg ala arg ala ala
61/21	91/31
ACG ACG TCG ATG GCC GCG CCC CGG CGG CTG	CAG CTG CGT AGC TCG ACC CGG TCG ACG ACG
thr thr ser met ala ala pro arg arg leu	gln leu arg ser ser thr arg ser thr thr
121/41	151/51
ACG GGG TCG GCG GGC CAG TCG GCG ATG TCG	AGG CGA TGG CAA TAC AGC GCC TTG GTG CGC
thr gly ser ala gly gln ser ala met ser	arg arg trp gln tyr ser ala leu val arg
181/61	211/71
GGC CAC ACG TCT GAG GTG GCG AAG ACC AGT	CCC GCG CCC ACC GGC AGC CGG ATC
gly his thr ser glu val ala lys thr ser	pro ala pro thr gly ser arg ile

SEQ ID N° 15C'

52/185

ORF contenant Seq15A' d'après Cole et al. (Nature 393:537-544)

```

1/1                                31/11
taa ggt ccg cca acg ctt tac gct cga cgg ccg cca cga gtt ggc cgg cca ctt tca ggc
OCH gly pro pro thr leu tyr ala arg arg pro pro arg val gly arg pro leu ser gly
61/21                                91/31
cgt agt cgc cgc agg gca ggg ctt ccc gcg tcg tct tcg cgg gtt tgt cgg caa agg tgt
arg ser arg arg arg ala gly leu pro ala ser ser ser arg val cys arg gln arg cys
121/41                                151/51
agg ggt agc gtt cgt ggg cgt cga cga cga tgt gca gct cgg gga tgc cgg cgg cgc ggg
arg gly ser val arg gly arg arg arg arg cys ala ala arg gly cys arg arg arg gly
181/61                                211/71
cgg tgg ggg tgc gca cgc ccg gcc gcg act gtt tgc gcg ttt tgg ggc tct gcc aga aca
arg trp gly cys ala arg pro ala ala thr val cys ala phe trp gly ser ala arg thr
241/81                                271/91
cca cct ggc ggc cgc gcg cca tgg tgt gca cca gtt gcg atc ggt tct ccc gcg cgc ggg
pro pro gly gly arg ala pro trp cys ala pro val ala ile gly ser pro ala arg gly
301/101                                331/111
cgg cga cga cgt cga tgg ccg cgc ccc ggc ggc tgc agc tgc gta gct cga ccc ggt cga
arg arg arg arg arg trp pro arg pro gly gly cys ser cys val ala arg pro gly arg
361/121                                391/131
cga cga cgg ggt cgg cgg gcc agt cgg cga tgt cga ggc gat ggc aat aca gcg cct tgg
arg arg arg gly arg arg ala ser arg arg cys arg gly asp gly asn thr ala pro trp
421/141                                451/151
tgc gcg gcc aca cgt ctg agg tgg cga aga cca gtc ccg cgc cca ccg gca gcc gga tca
cys ala ala thr arg leu arg trp arg arg pro val pro arg pro pro ala ala gly ser
481/161                                511/171
ggt agg gca ggc gcg agt ctt cag cgg ggt tgg cgg cga cga gca gct cca cag agt gtg
gly arg ala gly ala ser leu gln arg gly trp arg arg arg ala ala pro gln ser val
541/181                                571/191
agg gta cgg gcg gcg tac ggc aac ggt gaa gca ggc act ccg acg aac cca tcg tca cgt
arg val arg ala ala tyr gly asn gly glu ala gly thr pro thr asn pro ser ser arg
601/201
cga agg ggc agg tga
arg arg gly arg OPA

```

SEQ ID N° 15F

FIGURE 15F

53/185

R:Rv2530c prédite d'après Cole et al. (Nature 393:537-544) et pouvant être en phase avec SEQ15A

```

1/1                               31/11
gtg acg gca ctg ctc gat gtc aat gtg ctg atc gcg ctg ggc tgg ccg aat cac gtt cac
val thr ala leu leu asp val asn val leu ile ala leu gly trp pro asn his val his
61/21                               91/31
cat gcg gcc gcg cag cga tgg ttc acg cag ttc tcc tcg aat ggg tgg gcc acc acg ccg
his ala ala ala gln arg trp phe thr gln phe ser ser asn gly trp ala thr thr pro
121/41                               151/51
atc acc gag gca ggg tat gtc cga att tca agc aat cgc agt gtg atg cag gtg tcg acc
ile thr glu ala gly tyr val arg ile ser ser asn arg ser val met gln val ser thr
181/61                               211/71
acg ccg gct atc gcg atc gct cag ttg gcg gcg atg act tct ctt gcc ggg cac acg ttt
thr pro ala ile ala ile ala gln leu ala ala met thr ser leu ala gly his thr phe
241/81                               271/91
tgg cct gac gat gtg cca ctg atc gtt ggg agc gcc ggc gat cgc gat gcg gtg tcc aac
trp pro asp asp val pro leu ile val gly ser ala gly asp arg asp ala val ser asn
301/101                               331/111
cac cgt cgg gtc acc gac tgc cat ctc atc gcc ttg gcc gcg cgc tac ggg ggc cgg ttg
his arg arg val thr asp cys his leu ile ala leu ala ala arg tyr gly gly arg leu
361/121                               391/131
gtc aca ttc gat gcc gca ctg gcc gat tca gca tcc gca ggc ctc gtc gag gtg ttg tag
val thr phe asp ala ala leu ala asp ser ala ser ala gly leu val glu val leu AMB

```

SEQ ID N° 15R

FIGURE 15R

Seq15P: ORF d'après Cole et al. (Nature 393:537-544) contenant Rv2530c

```

1/1                               31/11
tga tgt tcc gcc gga tgc gcc gac ggt gac ttc cga gga tgt cgt ccg cgc gct cga gga
OPA cys ser ala gly cys ala asp gly asp phe arg gly cys arg pro arg ala arg gly
61/21                               91/31
cga cgt gtg acg gca ctg ctc gat gtc aat gtg ctg atc gcg ctg ggc tgg ccg aat cac
arg arg val thr ala leu leu asp val asn val leu ile ala leu gly trp pro asn his
121/41                               151/51
gtt cac cat gcg gcc gcg cag cga tgg ttc acg cag ttc tcc tcg aat ggg tgg gcc acc
val his his ala ala ala gln arg trp phe thr gln phe ser ser asn gly trp ala thr
181/61                               211/71
acg ccg atc acc gag gca ggg tat gtc cga att tca agc aat cgc agt gtg atg cag gtg
thr pro ile thr glu ala gly tyr val arg ile ser ser asn arg ser val met gln val
241/81                               271/91
tcg acc acg ccg gct atc gcg atc gct cag ttg gcg gcg atg act tct ctt gcc ggg cac
ser thr thr pro ala ile ala ile ala gln leu ala ala met thr ser leu ala gly his
301/101                               331/111
acg ttt tgg cct gac gat gtg cca ctg atc gtt ggg agc gcc ggc gat cgc gat gcg gtg
thr phe trp pro asp asp val pro leu ile val gly ser ala gly asp arg asp ala val
361/121                               391/131
tcc aac cac cgt cgg gtc acc gac tgc cat ctc atc gcc ttg gcc gcg cgc tac ggg ggc
ser asn his arg arg val thr asp cys his leu ile ala leu ala ala arg tyr gly gly
421/141                               451/151
cgg ttg gtc aca ttc gat gcc gca ctg gcc gat tca gca tcc gca ggc ctc gtc gag gtg
arg leu val thr phe asp ala ala leu ala asp ser ala ser ala gly leu val glu val
481/161
ttg tag
leu AMB

```

SEQ ID N° 15P

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 15P

54/185

Fragment contenant Seq15P' et Seq 15F'

1/1 31/11
tga tgt tcc gcc gga tgc gcc gac ggt gac ttc cga gga tgt cgt ccg cgc gct cga gga
OPA cys ser ala gly cys ala asp gly asp phe arg gly cys arg pro arg ala arg gly
asp val pro pro asp ala pro thr val thr ser glu asp val val arg ala leu glu asp
Met phe arg arg met arg arg arg OPA leu pro arg met ser ser ala arg ser arg thr
61/21 91/31
cga cgt gtg acg gca ctg ctc gat gtc aat gtg ctg atc gcg ctg ggc tgg ccg aat cac
arg arg val thr ala leu leu asp val asn val leu ile ala leu gly trp pro asn his
asp val OPA arg his cys ser met ser met cys OPA ser arg trp ala gly arg ile thr
thr cys asp gly thr ala arg cys gln cys ala asp arg ala gly leu ala glu ser arg
121/41 151/51
gtt cac cat gcg gcc gcg cag cga tgg ttc acg cag ttc tcc tcg aat ggg tgg gcc acc
val his his ala ala ala gln arg trp phe thr gln phe ser ser asn gly trp ala thr
phe thr met arg pro arg ser asp gly ser arg ser ser pro arg met gly gly pro pro
ser pro cys gly arg ala ala met val his ala val leu leu glu trp val gly his his
181/61 211/71
acg ccg atc acc gag gca ggg tat gtc cga att tca agc aat cgc agt gtg atg cag gtg
thr pro ile thr glu ala gly tyr val arg ile ser ser asn arg ser val met gln val
arg arg ser pro arg gln gly met ser glu phe gln ala ile ala val OPA cys arg cys
ala asp his arg gly arg val cys pro asn phe lys gln ser gln cys asp ala gly val
241/81 271/91
tcg acc acg ccg gct atc gcg atc gct cag ttg gcg gcg atg act tct ctt gcc ggg cac
ser thr thr pro ala ile ala ile ala gln leu ala ala met thr ser leu ala gly his
arg pro arg arg leu ser arg ser leu ser trp arg arg OPA leu leu leu pro gly thr
asp his ala gly tyr arg asp arg ser val gly gly asp phe ser cys arg ala his
301/101 331/111
acg ttt tgg cct gac gat gtg cca ctg atc gtt ggg agc gcc ggc gat cgc gat gcg gtg
thr phe trp pro asp asp val pro leu ile val gly ser ala gly asp arg asp ala val
arg phe gly leu thr met cys his OPA ser leu gly ala pro ala ile ala met arg cys
val leu ala OPA arg cys ala thr asp arg trp glu arg arg ser arg cys gly val
361/121 391/131
tcc aac cac cgt cgg gtc acc gac tgc cat ctc atc gcc ttg gcc gcg cgc tac ggg ggc
ser asn his arg arg val thr asp cys his leu ile ala leu ala ala arg tyr gly gly
pro thr thr val gly ser pro thr ala ile ser ser pro trp pro arg ala thr gly ala
gln pro pro ser gly his arg leu pro ser his arg leu gly arg ala leu arg gly pro
421/141 451/151
cgg ttg gtc aca ttc gat gcc gca ctg gcc gat tca gca tcc gca ggc ctc gtc gag gtg
arg leu val thr phe asp ala ala leu ala asp ser ala ser ala gly leu val glu val
gly trp ser his ser met pro his trp pro ile gln his pro gln ala ser ser arg cys
val gly his ile arg cys arg thr gly arg phe ser ile arg arg pro arg arg gly val
481/161 511/171
ttg tag tca ccg ggg atg ggc ggc tcg cca ggc ctg cag gat ctg cgg gcg cag gcg ccc
leu AMB ser pro gly met gly gly ser pro gly leu gln asp leu arg ala gln ala pro
cys ser his arg gly trp ala ala arg gln ala cys arg ile cys gly arg arg pro
val val thr gly asp gly arg leu ala arg pro ala gly ser ala gly ala gly ala pro
541/181 571/191
ccg gtc gga cac cgg cag gcc gac gct ttt ggc cca cgc gcg cag ctc ggc gct gct ggg
pro val gly his arg gln ala asp ala phe gly pro arg ala gln leu gly ala ala gly
arg ser asp thr gly arg pro thr leu leu ala his ala arg ser ser ala leu leu gly
gly arg thr pro ala gly arg arg phe trp pro thr arg ala ala arg arg cys trp ala
601/201 631/211
ctc ggg ctc ggc ggc agc cgg ctc gaa aac cgt ggt ggc gtc ggc atc gtc gac gaa cca
leu gly leu gly gly ser arg leu glu asn arg gly gly val gly ile val asp glu pro
ser gly ser ala ala ala gly ser lys thr val val ala ser ala ser ser thr asn gln
arg ala arg arg gln pro ala arg lys pro trp trp arg arg his arg arg arg thr arg

SEQ ID N° 15Q

FIGURE 15Q

FEUILLE DE REMPLACEMENT (REGLE 26)

55/185

661/221 691/231
 ggt gag ggc ggc ggc tag ata gcg gta ggt gta ttc ctg ggc gag ctt gcg ggt ttg gca
 gly glu gly gly gly AMB ile ala val gly val phe leu gly glu leu ala gly leu ala
 val arg ala ala ala arg AMB arg AMB val tyr ser trp ala ser leu arg val trp gln
 OPA gly arg arg leu asp ser gly arg cys ile pro gly arg ala cys gly phe gly arg
 721/241 751/251
 gaa cac gat cgg cac gtt ggg aaa gcc gat ctg caa ttc ggc cag ccc atc ggc gat cgc
 glu his asp arg his val gly lys ala asp leu gln phe gly gln pro ile gly asp arg
 asn thr ile gly thr leu gly lys pro ile cys asn ser ala ser pro ser ala ile ala
 thr arg ser ala arg trp glu ser arg ser ala ile arg pro ala his arg arg ser pro
 781/261 811/271
 cgt cgg gcg ggc gaa gga gtg cgc gaa gat ctc cga gta gcg gtc ctc gac cac cac ggc
 arg arg ala gly glu gly val arg glu asp leu arg val ala val leu asp his his gly
 val gly arg ala lys glu cys ala lys ile ser glu AMB arg ser ser thr thr thr ala
 ser gly gly arg arg ser ala arg arg ser pro ser ser gly pro arg pro pro arg arg
 841/281 871/291
 ggc ccg tgg cag cgc ggc cag ttc ggt cag ttg gta ttt cag gtt gcc gtt cag cac gcc
 gly pro trp gln arg gly gln phe gly gln leu val phe gln val ala val gln his ala
 ala arg gly ser ala ala ser ser val ser trp tyr phe arg leu pro phe ser thr pro
 pro val ala ala arg pro val arg ser val gly ile ser gly cys arg ser ala arg gln
 901/301 931/311
 aga agt aag gtc cgc caa cgc ttt acg ctc gac ggc cgc cac gag ttg gcc ggc cac ttt
 arg ser lys val arg gln arg phe thr leu asp gly arg his glu leu ala gly his phe
 glu val arg ser ala asn ala leu arg ser thr ala ala thr ser trp pro ala thr phe
 lys OCH gly pro pro thr leu tyr ala arg arg pro pro arg val gly arg pro leu ser
 961/321 991/331
 cag gcc gta gtc gcc gca ggg cag ggc ttc ccg cgt cgt ctt cgc ggg ttt gtc ggc aaa
 gln ala val val ala ala gly gln gly phe pro arg arg leu arg gly phe val gly lys
 arg pro AMB ser pro gln gly arg ala ser arg val val phe ala gly leu ser ala lys
 gly arg ser arg arg arg ala gly leu pro ala ser ser ser arg val cys arg gln arg
 1021/341 1051/351
 ggt gta ggg gta gcg ttc gtg ggc gtc gac gac gat gtg cag ctc ggg gat gcc ggc ggc
 gly val gly val ala phe val gly val asp asp asp val gln leu gly asp ala gly gly
 val AMB gly AMB arg ser trp ala ser thr thr met cys ser ser gly met pro ala ala
 cys arg gly ser val arg gly arg arg arg arg cys ala ala arg gly cys arg arg arg
 1081/361 1111/371
 gcg ggc ggt ggg ggt gcg cac gcc cgg ccg cga ctg ttt gcg cgt ttt ggg gct ctg cca
 ala gly gly gly gly ala his ala arg pro arg leu phe ala arg phe gly ala leu pro
 arg ala val gly val arg thr pro gly arg asp cys leu arg val leu gly leu cys gln
 gly arg trp gly cys ala arg pro ala ala thr val cys ala phe trp gly ser ala arg
 1141/381 1171/391
 gaa cac cac ctg gcg gcc gcg cgc cat ggt gtg cac cag ttg cga tcg gtt ctc ccg cgc
 glu his his leu ala ala ala arg his gly val his gln leu arg ser val leu pro arg
 asn thr thr trp arg pro arg ala met val cys thr ser cys asp arg phe ser arg ala
 thr pro pro gly gly arg ala pro trp cys ala pro val ala ile gly ser pro ala arg
 1201/401 1231/411
 gcg ggc ggc gac gac gtc gat ggc cgc gcc ccg gcg gct gca gct gcg tag ctc gac ccg
 ala gly gly asp asp val asp gly arg ala pro ala ala ala ala ala AMB leu asp pro
 arg ala ala thr thr ser met ala ala pro arg arg leu gln leu arg ser ser thr arg
 gly arg arg arg arg arg trp pro arg pro gly gly cys ser cys val ala arg pro gly
 1261/421 1291/431
 gtc gac gac gac ggg gtc ggc ggg cca gtc ggc gat gtc gag gcg atg gca ata cag cgc
 val asp asp asp gly val gly gly pro val gly asp val glu ala met ala ile gln arg
 ser thr thr thr gly ser ala gly gln ser ala met ser arg arg trp gln tyr ser ala
 arg arg arg arg gly arg arg ala ser arg arg cys arg gly asp gly asn thr ala pro

SEQ ID N° 15Q (suite 1)

FIGURE 15Q (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

56/185

1321/441
 ctt ggt gcg cgg cca cac gtc tga ggt ggc gaa gac cag tcc cgc gcc cac cgg cag ccg
 leu gly ala arg pro his val OPA gly gly glu asp gln ser arg ala his arg gln pro
 leu val arg gly his thr ser glu val ala lys thr ser pro ala pro thr gly ser arg
 trp cys ala ala thr arg leu arg trp arg arg pro val pro arg pro pro ala ala gly
 1381/461
 gat cag gta ggg cag gcg cga gtc ttc agc ggg gtt ggc gcc gac gag cag ctc cac aga
 asp gln val gly gln ala arg val phe ser gly val gly gly asp glu gln leu his arg
 ile arg AMB gly arg arg glu ser ser ala gly leu ala ala thr ser ser ser thr glu
 ser gly arg ala gly ala ser leu gln arg gly trp arg arg arg ala ala pro gln ser
 1441/481
 gtg tga ggg tac ggg cgg cgt acg gca acg gtg aag cag gca ctc cga cga acc cat cgt
 val OPA gly tyr gly arg arg thr ala thr val lys gln ala leu arg arg thr his arg
 cys glu gly thr gly gly val arg gln arg OPA ser arg his ser asp glu pro ile val
 val arg val arg ala ala tyr gly asn gly glu ala gly thr pro thr asn pro ser ser
 1501/501
 cac gtc gaa ggg gca ggt ga
 his val glu gly ala gly
 thr ser lys gly gln val
 arg arg arg gly arg OPA

SEQ ID N° 15Q (suite 2)

FIGURE 15Q (suite (2))

31/11
 TGC GCA TGC CGA CCA GTG TGG TTG GCC GGA GTT CGT TTG TTC GCG ATT GCC TCA ACG ATT
 cys ala cys arg pro val trp leu ala gly val arg leu phe ala ile ala ser thr ile
 61/21
 CGA TAT AAC CAC TCT AGT CAC ATC AAC CAC ACT CGT ACC ATC GAG CGT GTG GGT TCA TGC
 arg tyr asn his ser ser his ile asn his thr arg thr ile glu arg val gly ser cys
 121/41
 CAT GCA TTC GCG ACC GCG GGA GCC GGC GAA CCC GGC GCC ACA CAT AAT CCA GAT TGA GGA
 his ala phe ala thr ala gly ala gly glu pro gly ala thr his asn pro asp OPA gly
 181/61
 GAC TTC CGT GCC GAA CCG ACG CCG ACG CAA GCT TTC GAC AGC CAT GAG CGC GGT CGC CGC
 asp phe arg ala glu pro thr pro thr gln ala phe asp ser his glu arg gly arg arg
 241/81
 CCT GGC AGT TGC AAG TCC TTG TGC ATA TTT TCT TGT CTA CGA ATC AAC CGA AAC GAC CGA
 pro gly ser cys lys ser leu cys ile phe ser cys leu arg ile asn arg asn asp arg
 301/101
 GCG GCC CGA GCA CCA TGA ATT CAA GCA GGC GGC GGT GTT GAC CGA CCT GCC CGG CGA GCT
 ala ala arg ala pro OPA ile gln ala gly gly gly val asp arg pro ala arg arg ala
 361/121
 GAT GTC CGC GCT ATC GCA GGG GTT GTC CCA GTT CGG GAT C
 asp val arg ala ile ala gly val val pro val arg asp

SEQ ID N° 16A

FIGURE 16A

57/185

32/11
 GCG CAT GCC GAC CAG TGT GGT TGG CCG GAG TTC GTT TGT TCG CGA TTG CCT CAA CGA TTC
 ala his ala asp gln cys gly trp pro glu phe val cys ser arg leu pro gln arg phe
 62/21
 GAT ATA ACC ACT CTA GTC ACA TCA ACC ACA CTC GTA CCA TCG AGC GTG TGG GTT CAT GCC
 asp ile thr thr leu val thr ser thr thr leu val pro ser ser val trp val his ala
 122/41
 ATG CAT TCG CGA CCG CGG GAG CCG GCG AAC CCG GCG CCA CAC ATA ATC CAG ATT GAG GAG
 met his ser arg pro arg glu pro ala asn pro ala pro his ile ile gln ile glu glu
 182/61
 ACT TCC GTG CCG AAC CGA CGC CGA CGC AAG CTT TCG ACA GCC ATG AGC GCG GTC GCC GCC
 thr ser val pro asn arg arg arg arg lys leu ser thr ala met ser ala val ala ala
 242/81
 CTG GCA GTT GGA AGT CCT TGT GCA TAT TTT CTT GTC TAC GAA TCA ACC GAA ACG ACC GAG
 leu ala val ala ser pro cys ala tyr phe leu val tyr glu ser thr glu thr thr glu
 302/101
 CGG CCC GAG CAC CAT GAA TTC AAG CAG GCG GCG GTG TTG ACC GAC CTG CCC GGC GAG CTG
 arg pro glu his his glu phe lys gln ala ala val leu thr asp leu pro gly glu leu
 362/121
 ATG TCC GCG CTA TCG CAG GGG TTG TCC CAG TCC GGG ATC
 met ser ala leu ser gln gly leu ser gln phe gly ile
 392/131

SEQ ID N° 16B

FIGURE 16B

33/11
 CGC ATG CCG ACC AGT GTG GTT GGC CGG AGT TCG TTT GTT CGC GAT TGC CTC AAC GAT TCG
 arg met pro thr ser val val gly arg ser ser phe val arg asp cys leu asn asp ser
 63/21
 ATA TAA CCA CTC TAG TCA CAT CAA CCA CAC TCG TAC CAT CGA GCG TGT GGG TTC ATG CCA
 ile OCH pro leu AMB ser his gln pro his ser tyr his arg ala cys gly phe met pro
 123/41
 TGC ATT CGC GAC CGC GGG AGC CGG CGA ACC CGG CGC CAC ACA TAA TCC AGA TTG AGG AGA
 cys ile arg asp arg gly ser arg arg thr arg arg his thr OCH ser arg leu arg arg
 183/61
 CTT CCG TGC CGA ACC GAC GCC GAC GCA AGC TTT CGA CAG CCA TGA GCG CGG TCG CCG CCC
 leu pro cys arg thr asp ala asp ala ser phe arg gln pro OPA ala arg ser pro pro
 243/81
 TGG CAG TTG CAA GTC CTT GTG CAT ATT TTC TTG TCT ACG AAT CAA CCG AAA CGA CCG AGC
 trp gln leu gln val leu val his ile phe leu ser thr asn gln pro lys arg pro ser
 303/101
 GGC CCG AGC ACC ATG AAT TCA AGC AGG CGG CGG TGT TGA CCG ACC TGC CCG GCG AGC TGA
 gly pro ser thr met asn ser ser arg arg arg cys OPA pro thr cys pro ala ser OPA
 363/121
 TGT CCG CGC TAT CGC AGG GGT TGT CCC AGT TCG GGA TC
 cys pro arg tyr arg arg gly cys pro ser ser gly
 393/131

SEQ ID N° 16C

FIGURE 16C

FEUILLE DE REMPLACEMENT (REGLE 26)

58/185

31/11
 GCG GGC CAC CGA TCA GTC GAT CGG GTG GTT TCC GCT CCA TCA GCC CGG AAT TGA GGT GCC
 ala gly his arg ser val asp arg val val ser ala pro ser ala arg asn OPA gly ala
 61/21
 GCA GTG ACG ACA CCA GCG CAG GAC GCG CCG TTG GTG TTT CCC TCT GTT GCT TTC CCG TCC
 ala val thr thr pro ala gln asp ala pro leu val phe pro ser val ala phe pro ser
 121/41
 GGC TCG CCT TTT TTT CAT CAA CGT TGG ACT GCC GCA GTG GCG ATG TTG GTC GCC GGC GTG
 gly ser pro phe phe his gln arg trp thr ala ala val ala met leu val ala gly val
 181/61
 TTC GGT CAC CTG ACG GTC GGG ATG TTC CTT GGG TCT CGG GTT GCT GCT GGG TTT GCT CAA
 phe gly his leu thr val gly met phe leu gly ser arg val ala ala gly phe ala gln
 241/81
 TGC CCT GCT GGT GCG GCG TTC GGC CGA GTC GAT CAC CGC CAA AGA GCA CCC GTT AAA ACG
 cys pro ala gly ala ala phe gly arg val asp his arg gln arg ala pro val lys thr
 301/101
 GTC GAT GGC CCT CAA CTC GGC ATC GCG ACT GGC GAT TAT CAC CAT GCC TCG GGC TGA TC
 val asp gly pro gln leu gly ile ala thr gly asp tyr his his ala ser gly OPA

SEQ ID N° 17A

FIGURE 17A

32/11
 CGG GCC ACC GAT CAG TCG ATC GGG TGG TTT CCG CTC CAT CAG CCC GGA ATT GAG GTG CCG
 arg ala thr asp gln ser ile gly trp phe pro leu his gln pro gly ile glu val pro
 62/21
 CAG TGA CGA CAC CAG CGC AGG ACG CGC CGT TGG TGT TTC CCT CTG TTG CTT TCC CGT CCG
 gln OPA arg his gln arg arg thr arg arg trp cys phe pro leu leu leu ser arg pro
 122/41
 GCT CGC CTT TTT TTC ATC AAC GTT GGA CTG CCG CAG TGG CGA TGT TGG TCG CCG GCG TGT
 ala arg leu phe phe ile asn val gly leu pro gln trp arg cys trp ser pro ala cys
 182/61
 TCG GTC ACC TGA CGG TCG GGA TGT TCC TTG GGT CTC GGG TTG CTG CTG GGT TTG CTC AAT
 ser val thr OPA arg ser gly cys ser leu gly leu gly leu leu leu gly leu leu asn
 242/81
 GCC CTG CTG GTG CGG CGT TCG GCC GAG TCG ATC ACC GCC AAA GAG CAC CCG TTA AAA CCG
 ala leu leu val arg arg ser ala glu ser ile thr ala lys glu his pro leu lys arg
 302/101
 TCG ATG GCC CTC AAC TCG GCA TCG CGA CTG GCG ATT ATC ACC ATG CCT CGG GCT GAT C
 ser met ala leu asn ser ala ser arg leu ala ile ile thr met pro arg ala asp

SEQ ID N° 17B

FIGURE 17B

59/185

33/11
 GGG CCA CCG ATC AGT CGA TCG GGT GGT TTC CGC TCC ATC AGC CCG GAA TTG AGG TGC CGC
 gly pro pro ile ser arg ser gly gly phe arg ser ile ser pro glu leu arg cys arg
 63/21
 AGT GAC GAC ACC AGC GCA GGA CGC GCC GTT GGT GTT TCC CTC TGT TGC TTT CCC GTC CGG
 ser asp asp thr ser ala gly arg ala val gly val ser leu cys cys phe pro val arg
 123/41
 CTC GCC TTT TTT TCA TCA ACG TTG GAC TGC CGC AGT GGC GAT GTT GGT CGC CGG CGT GTT
 leu ala phe phe ser ser thr leu asp cys arg ser gly asp val gly arg arg arg val
 183/61
 CGG TCA CCT GAC GGT CGG GAT GTT CCT TGG GTC TCG GGT TGC TGC TGG GTT TGC TCA ATG
 arg ser pro asp gly arg asp val pro trp val ser gly cys cys trp val cys ser met
 243/81
 CCC TGC TGG TGC GGC GTT CGG CCG AGT CGA TCA CCG CCA AAG AGC ACC CGT TAA AAC GGT
 pro cys trp cys gly val arg pro ser arg ser pro pro lys ser thr arg OCH asn gly
 303/101
 CGA TGG CCC TCA ACT CGG CAT CGC GAC TGG CGA TTA TCA CCA TGC CTC GGG CTG ATC
 arg trp pro ser thr arg his arg asp trp arg leu ser pro cys leu gly leu ile

SEQ ID N° 17C

FIGURE 17C

partie de la séquence nucléotidique de seq17A

1/1 31/11
 ggc tag aac ccc gaa gga gac ctc gcg ggt tgc cgg ccc ccg gcc cat cgg atg cgt atc
 gly AMB asn pro glu gly asp leu ala gly cys arg pro pro ala his arg met arg ile
 61/21
 cgg tcg cgc cga ttc acg acc gac ata ggg agc tac ccc ttg ggt gat tcc ggt gcg acg
 arg ser arg arg phe thr thr asp ile gly ser tyr pro leu gly asp ser gly ala thr
 121/41
 act gcg ata cgc tcg gcg ggc cac cga tca gtc gat cgg gtg gtt tcc gct cca tca gcc
 thr ala ile arg ser ala gly his arg ser val asp arg val val ser ala pro ser ala
 181/61
 cgg aat tga ggt gcc gca gtg acg aca cca ggc cag gac gcg ccg ttg gtg ttt ccc tct
 arg asn OPA gly ala ala val thr thr pro ala gln asp ala pro leu val phe pro ser
 241/81
 gtt gct ttc cgt ccg gtt cgc ctt ttt ttc atc aac gtt gga ctg gcc gca gtg gcg atg
 val ala phe arg pro val arg leu phe phe ile asn val gly leu ala ala val ala met
 301/101
 ttg gtc gcc ggc gtg ttc ggt cac ctg acg gtc ggg atg ttc ttg ggt ctc ggg ttg ctg
 leu val ala gly val phe gly his leu thr val gly met phe leu gly leu gly leu leu
 361/121
 ctg ggt ttg ctc aat gcc ctg ctg gtg cgg cgt tcg gcc gag tcg atc acc gcc aaa gag
 leu gly leu leu asn ala leu leu val arg arg ser ala glu ser ile thr ala lys glu
 421/141
 cac ccg tta aaa cgg tcg atg gcc ctc aac tcg gca tcg cga ctg gcg att atc acc atc
 his pro leu lys arg ser met ala leu asn ser ala ser arg leu ala ile ile thr ile
 481/161
 ctc ggg ctg atc
 leu gly leu ile

SEQ ID N° 17A'

60/185

```

1/1                               31/11
gct aga acc ccg aag gag acc tcg cgg gtt gcc ggc ccc cgg ccc atc gga tgc gta tcc
ala arg thr pro lys glu thr ser arg val ala gly pro arg pro ile gly cys val ser
61/21                               91/31
ggt cgc gcc gat tca cga ccg aca tag gga gct acc cct tgg gtg att ccg gtg cga cga
gly arg ala asp ser arg pro thr AMB gly ala thr pro trp val ile pro val arg arg
121/41                               151/51
ctg cga tac gct cgg cgg gcc acc gat cag tcg atc ggg tgg ttt ccg ctc cat cag ccc
leu arg tyr ala arg arg ala thr asp gln ser ile gly trp phe pro leu his gln pro
181/61                               211/71
gga att gag gtg ccg cag tga cga cac cag cgc agg acg cgc cgt tgg tgt ttc cct ctg
gly ile glu val pro gln OPA arg his gln arg arg thr arg arg trp cys phe pro leu
241/81                               271/91
ttg ctt tcc gtc cgg ttc gcc ttt ttt tca tca acg ttg gac tgg ccg cag tgg cga tgt
leu leu ser val arg phe ala phe phe ser ser thr leu asp trp pro gln trp arg cys
301/101                               331/111
tgg tcg ccg gcg tgt tcg gtc acc tga cgg tcg gga tgt tct tgg gtc tcg ggt tgc tgc
trp ser pro ala cys ser val thr OPA arg ser gly cys ser trp val ser gly cys cys
361/121                               391/131
tgg gtt tgc tca atg ccc tgc tgg tgc gcc gtt cgg ccg agt cga tca ccg cca aag agc
trp val cys ser met pro cys trp cys gly val arg pro ser arg ser pro pro lys ser
421/141                               451/151
acc cgt taa aac ggt cga tgg ccc tca act cgg cat cgc gac tgg cga tta tca cca tcc
thr arg OCH asn gly arg trp pro ser thr arg his arg asp trp arg leu ser pro ser
481/161
tcg ggc tga tc
ser gly OPA

```

SEQ ID N° 17B'

FIGURE 17B'

```

1/1                               31/11
cta gaa ccc cga agg aga cct cgc ggg ttg ccg gcc ccc ggc cca tcg gat gcg tat ccg
leu glu pro arg arg arg pro arg gly leu pro ala pro gly pro ser asp ala tyr pro
61/21                               91/31
gtc gcg ccg att cac gac cga cat agg gag cta ccc ctt ggg tga ttc ccg tgc gac gac
val ala pro ile his asp arg his arg glu leu pro leu gly OPA phe arg cys asp asp
121/41                               151/51
tgc gat acg ctc ggc ggg cca ccg atc agt cga tcg ggt ggt ttc cgc tcc atc agc ccg
cys asp thr leu gly gly pro pro ile ser arg ser gly gly phe arg ser ile ser pro
181/61                               211/71
gaa ttg agg tgc cgc agt gac gac acc agc gca gga cgc gcc gtt ggt gtt tcc ctc tgt
glu leu arg cys arg ser asp asp thr ser ala gly arg ala val gly val ser leu cys
241/81                               271/91
tgc ttt ccg tcc ggt tcg cct ttt ttt cat caa cgt tgg act ggc cgc agt ggc gat gtt
cys phe pro ser gly ser pro phe phe his gln arg trp thr gly arg ser gly asp val
301/101                               331/111
ggt cgc ccg cgt gtt ccg tca cct gac ggt cgg gat gtt ctt ggg tct ccg gtt gct gct
gly arg arg arg val arg ser pro asp gly arg asp val leu gly ser arg val ala ala
361/121                               391/131
ggg ttt gct caa tgc cct gct ggt gcg gcg ttc gcc cga gtc gat cac cgc caa aga gca
gly phe ala gln cys pro ala gly ala ala phe gly arg val asp his arg gln arg ala
421/141                               451/151
ccc gtt aaa acg gtc gat gcc cct caa ctc gcc atc gcg act ggc gat tat cac cat cct
pro val lys thr val asp gly pro gln leu gly ile ala thr gly asp tyr his his pro
481/161
cgg gct gat c
arg ala asp

```

SEQ ID N° 17C'

FEUILLE DE REMPLACEMENT (REGLE 26)

61/185

séquence Rv1303 prédite par Cole et al. (Nature 393:537-544) et contenant partiellement Seq17A'

```

1/1                                     31/11
atg acg aca cca gcg cag gac gcg ccg ttg gtg ttt ccc tct gtt gct ttc cgt ccg gtt
met thr thr pro ala gln asp ala pro leu val phe pro ser val ala phe arg pro val
61/21                                     91/31
cgc ctt ttt ttc atc aac gtt gga ctg gcc gca gtg gcg atg ttg gtc gcc gcc gtg ttc
arg leu phe phe ile asn val gly leu ala ala val ala met leu val ala gly val phe
121/41                                     151/51
ggg cac ctg acg gtc ggg atg ttc ttg ggt ctc ggg ttg ctg ctg ggt ttg ctc aat gcc
gly his leu thr val gly met phe leu gly leu gly leu leu leu gly leu leu asn ala
181/61                                     211/71
ctg ctg gtg ccg cgt tcg gcc gag tcg atc acc gcc aaa gag cac ccg tta aaa ccg tcg
leu leu val arg arg ser ala glu ser ile thr ala lys glu his pro leu lys arg ser
241/81                                     271/91
atg gcc ctc aac tcg gca tcg cga ctg gcg att atc acc atc ctc ggg ctg atc atc gcc
met ala leu asn ser ala ser arg leu ala ile ile thr ile leu gly leu ile ile ala
301/101                                    331/111
tac att ttc ccg ccc gct gga ttg gcc gtc gtg ttc ggg ctg gca ttc ttc cag gtg ctg
tyr ile phe arg pro ala gly leu gly val val phe gly leu ala phe phe gln val leu
361/121                                    391/131
ctg gtg gca acg acg gcc ctg ccg gtc ctg aag aag ctg cgc act gcg acc gag gaa ccg
leu val ala thr thr ala leu pro val leu lys lys leu arg thr ala thr glu glu pro
421/141                                    451/151
gtc gca act tat tct tcc aat gcc cag acc ggg gga tcg gaa gga agg agc gcc agc gat
val ala thr tyr ser ser asn gly gln thr gly gly ser glu gly arg ser ala ser asp
481/161
gac tga
asp OPA

```

SEQ ID N° 17D

FIGURE 17D

Orf d'après Cole et al. (Nature 393:537-544) et contenant Rv1303

```

1/1                                     31/11
tga ggt gcc gca gtg acg aca cca gcg cag gac gcg ccg ttg gtg ttt ccc tct gtt gct
OPA gly ala ala val thr thr pro ala gln asp ala pro leu val phe pro ser val ala
61/21                                     91/31
ttc cgt ccg gtt cgc ctt ttt ttc atc aac gtt gga ctg gcc gca gtg gcg atg ttg gtc
phe arg pro val arg leu phe phe ile asn val gly leu ala ala val ala met leu val
121/41                                     151/51
gcc gcc gtg ttc ggt cac ctg acg gtc ggg atg ttc ttg ggt ctc ggg ttg ctg ctg ggt
ala gly val phe gly his leu thr val gly met phe leu gly leu gly leu leu leu gly
181/61                                     211/71
ttg ctc aat gcc ctg ctg gtg ccg cgt tcg gcc gag tcg atc acc gcc aaa gag cac ccg
leu leu asn ala leu leu val arg arg ser ala glu ser ile thr ala lys glu his pro
241/81                                     271/91
tta aaa ccg tcg atg gcc ctc aac tcg gca tcg cga ctg gcg att atc acc atc ctc ggg
leu lys arg ser met ala leu asn ser ala ser arg leu ala ile ile thr ile leu gly
301/101                                    331/111
ctg atc atc gcc tac att ttc ccg ccc gct gga ttg gcc gtc gtg ttc ggg ctg gca ttc
leu ile ile ala tyr ile phe arg pro ala gly leu gly val val phe gly leu ala phe
361/121                                    391/131
ttc cag gtg ctg ctg gtg gca acg acg gcc ctg ccg gtc ctg aag aag ctg cgc act gcg
phe gln val leu leu val ala thr thr ala leu pro val leu lys lys leu arg thr ala
421/141                                    451/151
acc gag gaa ccg gtc gca act tat tct tcc aat gcc cag acc ggg gga tcg gaa gga agg
thr glu glu pro val ala thr tyr ser ser asn gly gln thr gly gly ser glu gly arg
481/161
agc gcc agc gat gac tga
ser ala ser asp asp OPA

```

SEQ ID N° 17F

FEUILLE DE REMPLACEMENT (REGLE 26)

62/185

31/11
 GTC GAA CAG GTA CGG AAG GCG CCG TCG GTC GCT CGG TCC GCT GGT ATC TCG TGT TCA GCC
 val glu gln val arg lys ala pro ser val ala arg ser ala gly ile ser cys ser ala
 61/21
 AGC CAG CGG CCG TTA ACG TGG CCG AAC AGG TCG TCT TGG GGT CGG GCA TCA GCG TCG ATG
 ser gln arg pro leu thr trp pro asn arg ser ser trp gly arg ala ser ala ser met
 121/41
 TGG CTC AGG TCG ATA CCC GAG GGG ATG GCA AGT GTC ACC CCG CCA TCC TTC CAC CTC TTT
 trp leu arg ser ile pro glu gly met ala ser val thr pro pro ser phe his leu phe
 181/61
 TCG GGT GCA ACG ATC GGG CCA TGC CTG ACG GGG AGC AGA GCC AGC CAC CGG CCC AAG AAG
 ser gly ala thr ile gly pro cys leu thr gly ser arg ala ser his arg pro lys lys
 241/81
 ATG CGG AAG ACG ACT CGC GGC CCG ACG CCG CGG AGG CCG CCG CGG CCG AAC CCA AAT CAT
 met arg lys thr thr arg gly pro thr pro arg arg pro pro arg pro asn pro asn his
 301/101
 CAG CCG GTC CCG ATG TTC TCG ACC TAC GGT ATC GCC TCG ACA CTA CTC GGC GTG CTA TCG
 gln pro val pro met phe ser thr tyr gly ile ala ser thr leu leu gly val leu ser
 361/121
 GTC GCC GCG GTC GTG CTG GGT GCG ATG ATC
 val ala ala val val leu gly ala met ile

SEQ ID N° 18A

FIGURE 18A

32/11
 TCG AAC AGG TAC GGA AGG CGC CGT CGG TCG CTC GGT CCG CTG GTA TCT CGT GTT CAG CCA
 ser asn arg tyr gly arg arg arg arg ser leu gly pro leu val ser arg val gln pro
 62/21
 GCC AGC GGC CGT TAA CGT GGC CGA ACA GGT CGT CTT GGG GTC GGG CAT CAG CGT CGA TGT
 ala ser gly arg OCH arg gly arg thr gly arg leu gly val gly his gln arg arg cys
 122/41
 GGC TCA GGT CGA TAC CCG AGG GGA TGG CAA GTG TCA CCC CGC CAT CCT TCC ACC TCT TTT
 gly ser gly arg tyr pro arg gly trp gln val ser pro arg his pro ser thr ser phe
 182/61
 CGG GTG CAA CGA TCG GGC CAT GCC TGA CGG GGA GCA GAG CCA GCC ACC GGC CCA AGA AGA
 arg val gln arg ser gly his ala OPA arg gly ala glu pro ala thr gly pro arg arg
 242/81
 TGC GGA AGA CGA CTC GCG GCC CGA CGC CGC GGA GGC CGC CGC GGC CGA ACC CAA ATC ATC
 cys gly arg arg leu ala ala arg arg arg gly gly arg arg gly arg thr gln ile ile
 302/101
 AGC CGG TCC CGA TGT TCT CGA CCT ACG GTA TCG CCT CGA CAC TAC TCG GCG TGC TAT CGG
 ser arg ser arg cys ser arg pro thr val ser pro arg his tyr ser ala cys tyr arg
 362/121
 TCG CCG CGG TCG TGC TGG GTG CGA TGA TC
 ser pro arg ser cys trp val arg OPA

SEQ ID N° 18B

FIGURE 18B

FEUILLE DE REMPLACEMENT (REGLE 26)

63/185

3/1 33/11
 CGA ACA GGT ACG GAA GGC GCC GTC GGT CGC TCG GTC CGC TGG TAT CTC GTG TTC AGC CAG
 arg thr gly thr glu gly ala val gly arg ser val arg trp tyr leu val phe ser gln
 63/21 93/31
 CCA GCG GCC GTT AAC GTG GCC GAA CAG GTC GTC TTG GGG TCG GGC ATC AGC GTC GAT GTG
 pro ala ala val asn val ala glu gln val val leu gly ser gly ile ser val asp val
 123/41 153/51
 GCT CAG GTC GAT ACC CGA GGG GAT GGC AAG TGT CAC CCC GCC ATC CTT CCA CCT CTT TTC
 ala gln val asp thr arg gly asp gly lys cys his pro ala ile leu pro pro leu phe
 183/61 213/71
 GGG TGC AAC GAT CGG GCC ATG CCT GAC GGG GAG CAG AGC CAG CCA CCG GCC CAA GAA GAT
 gly cys asn asp arg ala met pro asp gly glu gln ser gln pro pro ala gln glu asp
 243/81 273/91
 GCG GAA GAC GAC TCG CGG CCC GAC GCC GCG GAG GCC GCC GCG GCC GAA CCC AAA TCA TCA
 ala glu asp asp ser arg pro asp ala ala glu ala ala ala glu pro lys ser ser
 303/101 333/111
 GCC GGT CCC GAT GTT CTC GAC CTA CGG TAT CGC CTC GAC ACT ACT CGG CGT GCT ATC GGT
 ala gly pro asp val leu asp leu arg tyr arg leu asp thr thr arg arg ala ile gly
 363/121
 CGC CGC GGT CGT GCT GGG TGC GAT GAT C
 arg arg gly arg ala gly cys asp asp

SEQ ID N° 18C

FIGURE 18C

partie de la séquence nucléotidique de seq18A

1/1 31/11
 GAA GGC GCC GTC GGT CGC TCG GTC CGC TGG TAT CTC GTG TTC AGC CAG CCA GCG GCC GTT
 glu gly ala val gly arg ser val arg trp tyr leu val phe ser gln pro ala ala val
 61/21 91/31
 AAC GTG GCC GAA CAG GTC GTC TTG GGG TCG GGC ATC AGC GTC GAT GTG GCT CAG GTC GAT
 asn val ala glu gln val val leu gly ser gly ile ser val asp val ala gln val asp
 121/41 151/51
 ACC CGA GGG GAT GGC AAG TGT CAC CCC GCC ATC CTT CCA CCT CTT TTC GGG TGC AAC GAT
 thr arg gly asp gly lys cys his pro ala ile leu pro pro leu phe gly cys asn asp
 181/61 211/71
 CGG GCC ATG CCT GAC GGG GAG CAG AGC CAG CCA CCG GCC CAA GAA GAT GCG GAA GAC GAC
 arg ala met pro asp gly glu gln ser gln pro pro ala gln glu asp ala glu asp asp
 241/81 271/91
 TCG CGG CCC GAC GCC GCG GAG GCC GCC GCG GCC GAA CCC AAA TCA TCA GCC GGT CCG ATG
 ser arg pro asp ala ala glu ala ala ala ala glu pro lys ser ser ala gly pro met
 301/101 331/111
 TTC TCG ACC TAC GGT ATC GCC TCG ACA CTA CTC GGC GTG CTA TCG GTC GCC GCG GTC GTG
 phe ser thr tyr gly ile ala ser thr leu leu gly val leu ser val ala ala val val
 361/121
 CTG GGT GCG ATG ATC
 leu gly ala met ile

SEQ ID N° 18A'

FIGURE 18A'

FEUILLE DE REMPLACEMENT (REGLE 26)

64/185

1/1 31/11
 CGG AAG GCG CCG TCG GTC GCT CGG TCC GCT GGT ATC TCG TGT TCA GCC AGC CAG CGG CCG
 arg lys ala pro ser val ala arg ser ala gly ile ser cys ser ala ser gln arg pro
 61/21 91/31
 TTA ACG TGG CCG AAC AGG TCG TCT TGG GGT CGG GCA TCA GCG TCG ATG TGG CTC AGG TCG
 leu thr trp pro asn arg ser ser trp gly arg ala ser ala ser met trp leu arg ser
 121/41 151/51
 ATA CCC GAG GGG ATG GCA AGT GTC ACC CCG CCA TCC TTC CAC CTC TTT TCG GGT GCA ACG
 ile pro glu gly met ala ser val thr pro pro ser phe his leu phe ser gly ala thr
 181/61 211/71
 ATC GGG CCA TGC CTG ACG GGG AGC AGA GCC AGC CAC CGG CCC AAG AAG ATG CGG AAG ACG
 ile gly pro cys leu thr gly ser arg ala ser his arg pro lys lys met arg lys thr
 241/81 271/91
 ACT CGC GGC CCG ACG CCG CGG AGG CCG CCG CGG CCG AAC CCA AAT CAT CAG CCG GTC CGA
 thr arg gly pro thr pro arg arg pro pro arg pro asn pro asn his gln pro val arg
 301/101 331/111
 TGT TCT CGA CCT ACG GTA TCG CCT CGA CAC TAC TCG GCG TGC TAT CGG TCG CCG CGG TCG
 cys ser arg pro thr val ser pro arg his tyr ser ala cys tyr arg ser pro arg ser
 361/121
 TGC TGG GTG CGA TGA TC
 cys trp val arg OPA

SEQ ID N° 18B'

FIGURE 18B'

1/1 31/11
 GGA AGG CGC CGT CGG TCG CTC GGT CCG CTG GTA TCT CGT GTT CAG CCA GCC AGC GGC CGT
 gly arg arg arg arg ser leu gly pro leu val ser arg val gln pro ala ser gly arg
 61/21 91/31
 TAA CGT GGC CGA ACA GGT CGT CTT GGG GTC GGG CAT CAG CGT CGA TGT GGC TCA GGT CGA
 OCH arg gly arg thr gly arg leu gly val gly his gln arg arg cys gly ser gly arg
 121/41 151/51
 TAC CCG AGG GGA TGG CAA GTG TCA CCC CGC CAT CCT TCC ACC TCT TTT CGG GTG CAA CGA
 tyr pro arg gly trp gln val ser pro arg his pro ser thr ser phe arg val gln arg
 181/61 211/71
 TCG GGC CAT GCC TGA CGG GGA GCA GAG CCA GCC ACC GGC CCA AGA AGA TGC GGA AGA CGA
 ser gly his ala OPA arg gly ala glu pro ala thr gly pro arg arg cys gly arg arg
 241/81 271/91
 CTC GCG GCC CGA CGC CGC GGA GGC CGC CGC GGC CGA ACC CAA ATC ATC AGC CGG TCC GAT
 leu ala ala arg arg arg gly gly arg arg gly arg thr gln ile ile ser arg ser asp
 301/101 331/111
 GTT CTC GAC CTA CGG TAT CGC CTC GAC ACT ACT CGG CGT GCT ATC GGT CGC CGC GGT CGT
 val leu asp leu arg tyr arg leu asp thr thr arg arg ala ile gly arg arg gly arg
 361/121
 GCT GGG TGC GAT GAT C
 ala gly cys asp asp

SEQ ID N° 18C'

FIGURE 18C'

FEUILLE DE REMPLACEMENT (REGLE 26)

65/185

séquence Rv0199 prédite par Cole et al. (Nature 393:537-544) et contenant seq18A'

1/1	31/11
atg cct gac ggg gag cag agc cag cca ccg gcc caa gaa gat gcg gaa gac gac tcg cgg	
Met pro asp gly glu gln ser gln pro pro ala gln glu asp ala glu asp asp ser arg	
61/21	91/31
ccc gac gcc gcg gag gcc gcc gcg gcc gaa ccc aaa tca tca gcc ggt ccg atg ttc tcg	
pro asp ala ala glu ala ala ala ala glu pro lys ser ser ala gly pro met phe ser	
121/41	151/51
acc tac ggt atc gcc tcg aca cta ctc ggc gtg cta tcg gtc gcc gcg gtc gtg ctg ggt	
thr tyr gly ile ala ser thr leu leu gly val leu ser val ala ala val val leu gly	
181/61	211/71
gcg atg atc tgg tcc gca cac cgc gat gac tcc ggc gag cgt acc tac ctg acc cgg gtc	
ala met ile trp ser ala his arg asp asp ser gly glu arg thr tyr leu thr arg val	
241/81	271/91
atg ctg acc gcc gct gaa tgg acg gcc gtg ctg atc aac atg aac gcc gac aac atc gat	
met leu thr ala ala glu trp thr ala val leu ile asn met asn ala asp asn ile asp	
301/101	331/111
gcc agc ctg cag cga ctg cac gac gga acg gtc ggt caa ctc aac acc gac ttc gac gct	
ala ser leu gln arg leu his asp gly thr val gly gln leu asn thr asp phe asp ala	
361/121	391/131
gtc gtg cag ccc tac cgg cag gtg gtg gag aag ttg cgg acg cac agc agc ggc agg atc	
val val gln pro tyr arg gln val val glu lys leu arg thr his ser ser gly arg ile	
421/141	451/151
gag gcg gta gcg atc gat acg gtg cac cgc gag ctg gat acc cag tcc ggt gcc gcc cga	
glu ala val ala ile asp thr val his arg glu leu asp thr gln ser gly ala ala arg	
481/161	511/171
ccg gta gta acc acg aaa ttg cca ccg ttt gcc act cgc acc gac tcg gtg ctg ctg gtc	
pro val val thr thr lys leu pro pro phe ala thr arg thr asp ser val leu leu val	
541/181	571/191
gcg acg tcg gtc agt gag aac gcc ggc gcc aaa ccc cag acc gtg cac tgg aac ttg cgg	
ala thr ser val ser glu asn ala gly ala lys pro gln thr val his trp asn leu arg	
601/201	631/211
ctc gat gtc tcc gat gtg gac ggc aag ctg atg atc tcc cgg ttg gag tcg att cga tga	
leu asp val ser asp val asp gly lys leu met ile ser arg leu glu ser ile arg OPA	

SEQ ID N° 18D

FIGURE 18D

66/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv0199

```

1/1
taa tcc gat gcc gga ttg ggt gaa atg cac
OCH ser asp ala gly leu gly glu met his
61/21
atc gac ata gac tcc gat gcc gcc gcc cac
ile asp ile asp ser asp ala ala ala his
121/41
ggc caa ttc ggt ggc gtc ggc cgc gct gtc
gly gln phe gly gly val gly arg ala val
181/61
cac ccc tgc gcg ctc gac ggc ttc ctc gtc
his pro cys ala leu asp gly phe leu val
241/81
ctg cgc atc ggt gcc tac cgc agc acc tgc
leu arg ile gly ala tyr arg ser thr cys
301/101
ttg tgt ctc ggc gcg gtc gaa cag gct acg
leu cys leu gly ala val glu gln ala thr
361/121
tat ctc gtg ttc agc cag cca gcg gcc gtt
tyr leu val phe ser gln pro ala ala val
421/141
ggc atc agc gtc gat gtg gct cag gtc gat
gly ile ser val asp val ala gln val asp
481/161
atc ctt cca cct ctt ttc ggg tgc aac gat
ile leu pro pro leu phe gly cys asn asp
541/181
cca ccg gcc caa gaa gat gcg gaa gac gac
pro pro ala gln glu asp ala glu asp asp
601/201
gcc gaa ccc aaa tca tca gcc ggt ccg atg
ala glu pro lys ser ser ala gly pro met
661/221
ctc ggc gtg cta tcg gtc gcc gcg gtc gtg
leu gly val leu ser val ala ala val val
721/241
gat gac tcc ggc gag cgt acc tac ctg acc
asp asp ser gly glu arg thr tyr leu thr
781/261
gcc gtg ctg atc aac atg aac gcc gac aac
ala val leu ile asn met asn ala asp asn
841/281
gga acg gtc ggt caa ctc aac acc gac ttc
gly thr val gly gln leu asn thr asp phe
901/301
gtg gag aag ttg ccg acg cac agc agc ggc
val glu lys leu arg thr his ser ser gly
961/321
cac cgc gag ctg gat acc cag tcc ggt gcc
his arg glu leu asp thr gln ser gly ala
1021/341
ccg ttt gcc act cgc acc gac tcg gtg ctg
pro phe ala thr arg thr asp ser val leu
1081/361
ggc gcc aaa ccc cag acc gtg cac tgg aac
gly ala lys pro gln thr val his trp asn
1141/381
aag ctg atg atc tcc ccg ttg gag tcg att
lys leu met ile ser arg leu glu ser ile
31/11
caa gta acg ggt cga gtc ttt gga atc ggt
gln val thr gly arg val phe gly ile gly
91/31
gcc ggc acg ttg cag agt gcc aag ggc ggc
ala gly thr leu gln ser ala lys gly gly
151/51
aat cgt ggc caa ttc gtc gtg cag ccg ttg
asn arg gly gln phe val val gln arg leu
211/71
gag gaa gct ggc gta gag gtc gcc gat gcg
glu glu ala gly val glu val ala asp ala
271/91
ttg gct ggc ctg gat gat cag gtc tcg cac
leu ala gly leu asp asp gln val ser his
331/111
gaa ggc gcc gtc ggt cgc tcg gtc cgc tgg
glu gly ala val gly arg ser val arg trp
391/131
aac gtg gcc gaa cag gtc gtc ttg ggg tcg
asn val ala glu gln val val leu gly ser
451/151
acc cga ggg gat ggc aag tgt cac ccc gcc
thr arg gly asp gly lys cys his pro ala
511/171
cgg gcc atg cct gac ggg gag cag agc cag
arg ala met pro asp gly glu gln ser gln
571/191
tcg ccg ccc gac gcc gcg gag gcc gcc gcg
ser arg pro asp ala ala glu ala ala ala
631/211
ttc tcg acc tac ggt atc gcc tcg aca cta
phe ser thr tyr gly ile ala ser thr leu
691/231
ctg ggt gcg atg atc tgg tcc gca cac cgc
leu gly ala met ile trp ser ala his arg
751/251
cgg gtc atg ctg acc gcc gct gaa tgg acg
arg val met leu thr ala ala glu trp thr
811/271
atc gat gcc agc ctg cag cga ctg cac gac
ile asp ala ser leu gln arg leu his asp
871/291
gac gct gtc gtg cag ccc tac ccg cag gtg
asp ala val val gln pro tyr arg gln val
931/311
agg atc gag gcg gta gcg atc gat acg gtg
arg ile glu ala val ala ile asp thr val
991/331
gcc cga ccg gta gta acc acg aaa ttg cca
ala arg pro val val thr thr lys leu pro
1051/351
ctg gtc gcg acg tcg gtc agt gag aac gcc
leu val ala thr ser val ser glu asn ala
1111/371
ttg ccg ctc gat gtc tcc gat gtg gac ggc
leu arg leu asp val ser asp val asp gly
1171/391
cga tga
OPA

```

SEQ ID N° 18F

FIGURE 18F
FEUILLE DE REMPLACEMENT (REGLE 26)

67/185

31/11
 GTT GCG CAA CGG GGT GAG CAC CGA CGC GAT GAT GGC GCA ACT ATC GAA ACT GCA GGA CAT
 val ala gln arg gly glu his arg arg asp asp gly ala thr ile glu thr ala gly his
 61/21
 CGC CAA CGC CAA CGA CGG CAC TCG CGC GGT GGG CAC CCC TGG CTA TCA GGC CAG CGT CGA
 arg gln arg gln arg arg his ser arg gly gly his pro trp leu ser gly gln arg arg
 121/41
 CTA TGT GGT AAA CAC ACT GCG CAA CAG CGG TTT TGA TGT GCA AAC CCC GGA GTT CTC CGC
 leu cys gly lys his thr ala gln gln arg phe OPA cys ala asn pro gly val leu arg
 181/61
 TCG CGT GTT CAA GGC CGA AAA AGG GGT GGT GAC CCT CGG CGG CAA CAC CGT GGA GGC GAG
 ser arg val gln gly arg lys arg gly gly asp pro arg arg gln his arg gly gly glu
 241/81
 GGC GCT CGA GTA CAG CCT CGG CAC ACC GCC GGA CGG GGT GAC GGG CCC GCT GGT GGC TGC
 gly ala arg val gln pro arg his thr ala gly arg gly asp gly pro ala gly gly cys
 301/101
 CCC CGC CGA CGA CAG TCC GGG CTG CAG TCC GTC GGA CTA CGA CAG GCT GCC GGT GTC CGG
 pro arg arg arg gln ser gly leu gln ser val gly leu arg gln ala ala gly val arg
 361/121
 TGC GGT GGT GCT GGT AGA TC
 cys gly gly ala gly arg

SEQ ID N° 19A

FIGURE 19A

32/11
 TTG CGC AAC GGG GTG AGC ACC GAC GCG ATG ATG GCG CAA CTA TCG AAA CTG CAG GAC ATC
 leu arg asn gly val ser thr asp ala met met ala gln leu ser lys leu gln asp ile
 62/21
 GCC AAC GCC AAC GAC GGC ACT CGC GCG GTG GGC ACC CCT GGC TAT CAG GCC AGC GTC GAC
 ala asn ala asn asp gly thr arg ala val gly thr pro gly tyr gln ala ser val asp
 122/41
 TAT GTG GTA AAC ACA CTG CGC AAC AGC GGT TTT GAT GTG CAA ACC CCG GAG TTC TCC GCT
 tyr val val asn thr leu arg asn ser gly phe asp val gln thr pro glu phe ser ala
 182/61
 CGC GTG TTC AAG GCC GAA AAA GGG GTG GTG ACC CTC GGC GGC AAC ACC GTG GAG GCG AGG
 arg val phe lys ala glu lys gly val val thr leu gly gly asn thr val glu ala arg
 242/81
 GCG CTC GAG TAC AGC CTC GGC ACA CCG CCG GAC GGG GTG ACG GGC CCG CTG GTG GCT GCC
 ala leu glu tyr ser leu gly thr pro pro asp gly val thr gly pro leu val ala ala
 302/101
 CCC GCC GAC GAC AGT CCG GGC TGC AGT CCG TCG GAC TAC GAC AGG CTG CCG GTG TCC GGT
 pro ala asp asp ser pro gly cys ser pro ser asp tyr asp arg leu pro val ser gly
 362/121
 GCG GTG GTG CTG GTA GAT C
 ala val val leu val asp

SEQ ID N° 19B

FIGURE 19B

68/185

33/11
TGC GCA ACG GGG TGA GCA CCG ACG CGA TGA TGG CGC AAC TAT CGA AAC TGC AGG ACA TCG
cys ala thr gly OPA ala pro thr arg OPA trp arg asn tyr arg asn cys arg thr ser
63/21
93/31
CCA ACG CCA ACG ACG GCA CTC GCG CGG TGG GCA CCC CTG GCT ATC AGG CCA GCG TCG ACT
pro thr pro thr thr ala leu ala arg trp ala pro leu ala ile arg pro ala ser thr
123/41
153/51
ATG TGG TAA ACA CAC TGC GCA ACA GCG GTT TTG ATG TGC AAA CCC CGG AGT TCT CCG CTC
met trp OCH thr his cys ala thr ala val leu met cys lys pro arg ser ser pro leu
183/61
213/71
GCG TGT TCA AGG CCG AAA AAG GGG TGG TGA CCC TCG GCG GCA ACA CCG TGG AGG CGA GGG
ala cys ser arg pro lys lys gly trp OPA pro ser ala ala thr pro trp arg arg gly
243/81
273/91
CGC TCG AGT ACA GCC TCG GCA CAC CGC CGG ACG GGG TGA CGG GCC CGC TGG TGG CTG CCC
arg ser ser thr ala ser ala his arg arg thr gly OPA arg ala arg trp trp leu pro
303/101
333/111
CCG CCG ACG ACA GTC CGG GCT GCA GTC CGT CGG ACT ACG ACA GGC TGC CGG TGT CCG GTG
pro pro thr thr val arg ala ala val arg arg thr thr thr gly cys arg cys pro val
363/121
CGG TGG TGC TGG TAG ATC
arg trp cys trp AMB ile

SEQ ID N° 19C

FIGURE 19C

partie de la séquence nucléotidique de seq19A

1/1 31/11
CTA TCG AAA CTG CAG GAC ATC GCC AAC GCC AAC GAC GGC ACT CGC GCG GTG GGC ACC CCT
leu ser lys leu gln asp ile ala asn ala asn asp gly thr arg ala val gly thr pro
61/21 91/31
GGC TAT CAG GCC AGC GTC GAC TAT GTG GTA AAC ACA CTG CGC AAC AGC GGT TTT GAT GTG
gly tyr gln ala ser val asp tyr val val asn thr leu arg asn ser gly phe asp val
121/41 151/51
CAA ACC CCG GAG TTC TCC GCT CGC GTG TTC AAG GCC GAA AAA GGG GTG GTG ACC CTC GGC
gln thr pro glu phe ser ala arg val phe lys ala glu lys gly val val thr leu gly
181/61 211/71
GGC AAC ACC GTG GAG GCG AGG GCG CTC GAG TAC AGC CTC GGC ACA CCG CCG GAC GGG GTG
gly asn thr val glu ala arg ala leu glu tyr ser leu gly thr pro pro asp gly val
241/81 271/91
ACG GGC CCG CTG GTG GCT GCC CCC GCC GAC GAC AGT CCG GGC TGC AGT CCG TCG GAC TAC
thr gly pro leu val ala ala pro ala asp asp ser pro gly cys ser pro ser asp tyr
301/101 331/111
GAC AGG CTG CCG GTG TCC GGT GCG GTG GTG CTG GTA GAT C
asp arg leu pro val ser gly ala val val leu val asp

SEQ ID N° 19A'

FIGURE 19A

69/185

```

1/1                               31/11
TAT CGA AAC TGC AGG ACA TCG CCA ACG CCA ACG ACG GCA CTC GCG CGG TGG GCA CCC CTG
tyr arg asn cys arg thr ser pro thr pro thr thr ala leu ala arg trp ala pro leu
61/21                               91/31
GCT ATC AGG CCA GCG TCG ACT ATG TGG TAA ACA CAC TGC GCA ACA GCG GTT TTG ATG TGC
ala ile arg pro ala ser thr met trp OCH thr his cys ala thr ala val leu met cys
121/41                               151/51
AAA CCC CGG AGT TCT CCG CTC GCG TGT TCA AGG CCG AAA AAG GGG TGG TGA CCC TCG GCG
lys pro arg ser ser pro leu ala cys ser arg pro lys lys gly trp OPA pro ser ala
181/61                               211/71
GCA ACA CCG TGG AGG CGA GGG CGC TCG AGT ACA GCC TCG GCA CAC CGC CGG ACG GGG TGA
ala thr pro trp arg arg gly arg ser ser thr ala ser ala his arg arg thr gly OPA
241/81                               271/91
CGG GCC CGC TGG TGG CTG CCC CCG CCG ACG ACA GTC CGG GCT GCA GTC CGT CGG ACT ACG
arg ala arg trp trp leu pro pro pro thr thr val arg ala ala val arg arg thr thr
301/101                               331/111
ACA GGC TGC CGG TGT CCG GTG CGG TGG TGC TGG TAG ATC
thr gly cys arg cys pro val arg trp cys trp AMB ile

```

SEQ ID N° 19B'

FIGURE 19B'

```

1/1                               31/11
ATC GAA ACT GCA GGA CAT CGC CAA CGC CAA CGA CGG CAC TCG CGC GGT GGG CAC CCC TGG
ile glu thr ala gly his arg gln arg gln arg arg his ser arg gly gly his pro trp
61/21                               91/31
CTA TCA GGC CAG CGT CGA CTA TGT GGT AAA CAC ACT GCG CAA CAG CGG TTT TGA TGT GCA
leu ser gly gln arg arg leu cys gly lys his thr ala gln gln arg phe OPA cys ala
121/41                               151/51
AAC CCC GGA GTT CTC CGC TCG CGT GTT CAA GGC CGA AAA AGG GGT GGT GAC CCT CGG CGG
asn pro gly val leu arg ser arg val gln gly arg lys arg gly gly asp pro arg arg
181/61                               211/71
CAA CAC CGT GGA GGC GAG GGC GCT CGA GTA CAG CCT CGG CAC ACC GCC GGA CGG GGT GAC
gln his arg gly gly glu gly ala arg val gln pro arg his thr ala gly arg gly asp
241/81                               271/91
GGG CCC GCT GGT GGC TGC CCC CGC CGA CGA CAG TCC GGG CTG CAG TCC GTC GGA CTA CGA
gly pro ala gly gly cys pro arg arg arg gln ser gly leu gln ser val gly leu arg
301/101                               331/111
CAG GCT GCC GGT GTC CGG TGC GGT GGT GCT GGT AGA TC
gln ala ala gly val arg cys gly gly ala gly arg

```

SEQ ID N° 19C'

FIGURE 19C'

70/185

sequence Rv0418 prédite par Cole et al. (Nature 393:537-544) et contenant seq19A'

1/1	31/11
atg gtg aac aaa tcc agg atg atg ccg gcg	gtg ctg gcc gtg gct gtg gtc gtc gca ttc
Met val asn lys ser arg met met pro ala	val leu ala val ala val val ala phe
61/21	91/31
ctg acg acg ggc tgt atc ccg tgg tct acg	cag tcg ccg ccc gtt gtt aac ggc ccc gct
leu thr thr gly cys ile arg trp ser thr	gln ser arg pro val val asn gly pro ala
121/41	151/51
gcc gca gag ttc gcc gtt gcg ttg cgc aac	cgg gtg agc acc gac gcg atg atg gcg cac
ala ala glu phe ala val ala leu arg asn	arg val ser thr asp ala met met ala his
181/61	211/71
cta tcg aaa ctg cag gac atc gcc aac gcc	aac gac ggc act cgc gcg gtg ggc acc cct
leu ser lys leu gln asp ile ala asn ala	asn asp gly thr arg ala val gly thr pro
241/81	271/91
ggc tat cag gcc agc gtc gac tat gtg gta	aac aca ctg cgc aac agc ggt ttt gat gtg
gly tyr gln ala ser val asp tyr val val	asn thr leu arg asn ser gly phe asp val
301/101	331/111
caa acc ccg gag ttc tcc gct cgc gtg ttc	aag gcc gaa aaa ggg gtg gtg acc ctc ggc
gln thr pro glu phe ser ala arg val phe	lys ala glu lys gly val val thr leu gly
361/121	391/131
ggc aac acc gtg gag gcg agg gcg ctc gag	tac agc ctc ggc aca ccg ccg gac ggg gtg
gly asn thr val glu ala arg ala leu glu	tyr ser leu gly thr pro pro asp gly val
421/141	451/151
acg ggc ccg ctg gtg gct gcc ccc gcc gac	gac agt ccg ggc tgc agt ccg tcg gac tac
thr gly pro leu val ala ala pro ala asp	asp ser pro gly cys ser pro ser asp tyr
481/161	511/171
gac agg ctg ccg gtg tcc ggt gcg gtg gtg	ctg gta gat cgc ggc gtc tgt cct ttt gcc
asp arg leu pro val ser gly ala val val	leu val asp arg gly val cys pro phe ala
541/181	571/191
cag aag gaa gac gca gcc gcg cag cgc ggt	gcg gtg gcg ctg atc att gct gac aac atc
gln lys glu asp ala ala ala gln arg gly	ala val ala leu ile ile ala asp asn ile
601/201	631/211
gac gag cag gcg atg ggc ggc acc ctg ggg	gct aat acc gac gtc aag atc ccg gtg gtg
asp glu gln ala met gly gly thr leu gly	ala asn thr asp val lys ile pro val val
661/221	691/231
agt gtc acc aag tcg gtc gga ttc cag cta	cgc gga cag tct ggg cca acc acc gtc aag
ser val thr lys ser val gly phe gln leu	arg gly gln ser gly pro thr thr val lys
721/241	751/251
ctc acg gcg agc acc caa agt ttc aag gcc	cgc aac gtc atc gcg cag acg aag acg ggg
leu thr ala ser thr gln ser phe lys ala	arg asn val ile ala gln thr lys thr gly
781/261	811/271
tcg tcg gcc aac gtg gtg atg gca ggt gcg	cat ttg gac agc gtt ccg gaa gga ccc ggc
ser ser ala asn val val met ala gly ala	his leu asp ser val pro glu gly pro gly
841/281	871/291
atc aac gac aac ggc tcg gga gtg gct gcg	gtt ctg gaa acg gca gtg cag ctg ggg aac
ile asn asp asn gly ser gly val ala ala	val leu glu thr ala val gln leu gly asn
901/301	931/311
tca ccg cat gtg tcc aac gcg gta ccg ttc	gcc ttc tgg ggc gcc gag gaa ttc ggc ctg
ser pro his val ser asn ala val arg phe	ala phe trp gly ala glu glu phe gly leu
961/321	991/331
att ggg tca cga aac tac gtc gag tcg ctg	gac atc gac gcg ctc aaa ggc atc gcg ctg
ile gly ser arg asn tyr val glu ser leu	asp ile asp ala leu lys gly ile ala leu

SEQ ID N° 19 D

FIGURE 19D

FEUILLE DE REMPLACEMENT (REGLE 26)

71/185

1021/341	1051/351
tat ctg aac ttc gac atg ttg gcg tcg ccg aac ccg ggt tac ttc acc tac gac ggt gac	
tyr leu asn phe asp met leu ala ser pro asn pro gly tyr phe thr tyr asp gly asp	
1081/361	1111/371
cag tcg ctg ccg cta gac gcc cgc ggt cag ccg gtg gtg ccc gaa ggc tcg gcc ggt atc	
gln ser leu pro leu asp ala arg gly gln pro val val pro glu gly ser ala gly ile	
1141/381	1171/391
gag cgc acg ttc gtc gcc tat ctg aag atg gcc ggc aag acc gcg cag gac acc tcg ttc	
glu arg thr phe val ala tyr leu lys met ala gly lys thr ala gln asp thr ser phe	
1201/401	1231/411
gac ggt cgg tcc gac tac gac ggc ttc acg ctg gcg ggt atc cct tcg ggt ggc ctg ttc	
asp gly arg ser asp tyr asp gly phe thr leu ala gly ile pro ser gly gly leu phe	
1261/421	1291/431
tcc ggc gct gag gtc aag aag tcc gcc gag caa gcc gag ctc tgg ggc ggc acc gcc gac	
ser gly ala glu val lys lys ser ala glu gln ala glu leu trp gly gly thr ala asp	
1321/441	1351/451
gag cct ttc gat ccc aac tat cac cag aag aca gac acc ctg gac cat atc gac cgc acc	
glu pro phe asp pro asn tyr his gln lys thr asp thr leu asp his ile asp arg thr	
1381/461	1411/471
gcg ctc ggt atc aac ggc gct ggc gtc gcg tac gcg gtg ggt ttg tat gcg cag gac ctc	
ala leu gly ile asn gly ala gly val ala tyr ala val gly leu tyr ala gln asp leu	
1441/481	1471/491
ggc ggc ccc aac ggg gtt ccg gtc atg gcg gac cgc acc cgc cac ctg att gcc aaa ccg	
gly gly pro asn gly val pro val met ala asp arg thr arg his leu ile ala lys pro	
1501/501	
tga	
OPA	

SEQ ID N° 19D (suite)

FIGURE 19D (suite)

72/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv0418

```

1/1                                31/11
tag gcc att caa cgc tct gtt cgt ttg att ggt cgg tgg gat gcg aaa gct gcg cgg cga
AMB ala ile gln arg ser val arg leu ile gly arg trp asp ala lys ala ala arg arg
61/21                                91/31
cag gcg cgg tct aat ctg ggc gcg atg gtg aac aaa tcc agg atg atg ccg gcg gtg ctg
gln ala arg ser asn leu gly ala met val asn lys ser arg met met pro ala val leu
121/41                                151/51
gcc gtg gct gtg gtc gtc gca ttc ctg acg acg ggc tgt atc cgg tgg tct acg cag tcg
ala val ala val val val ala phe leu thr thr gly cys ile arg trp ser thr gln ser
181/61                                211/71
cgg ccc gtt gtt aac ggc ccc gct gcc gca gag ttc gcc gtt gcg ttg cgc aac cgg gtg
arg pro val val asn gly pro ala ala ala glu phe ala val ala leu arg asn arg val
241/81                                271/91
agc acc gac gcg atg atg gcg cac cta tcg aaa ctg cag gac atc gcc aac gcc aac gac
ser thr asp ala met met ala his leu ser lys leu gln asp ile ala asn ala asn asp
301/101                               331/111
ggc act cgc gcg gtg ggc acc cct ggc tat cag gcc agc gtc gac tat gtg gta aac aca
gly thr arg ala val gly thr pro gly tyr gln ala ser val asp tyr val val asn thr
361/121                               391/131
ctg cgc aac agc ggt ttt gat gtg caa acc ccg gag ttc tcc gct cgc gtg ttc aag gcc
leu arg asn ser gly phe asp val gln thr pro glu phe ser ala arg val phe lys ala
421/141                               451/151
gaa aaa ggg gtg gtg acc ctc ggc ggc aac acc gtg gag gcg agg gcg ctc gag tac agc
glu lys gly val val thr leu gly gly asn thr val glu ala arg ala leu glu tyr ser
481/161                               511/171
ctc ggc aca ccg ccg gac ggg gtg acg ggc ccg ctg gtg gct gcc ccc gcc gac gac agt
leu gly thr pro pro asp gly val thr gly pro leu val ala ala pro ala asp asp ser
541/181                               571/191
ccg ggc tgc agt ccg tcg gac tac gac agg ctg ccg gtg tcc ggt gcg gtg gtg ctg gta
pro gly cys ser pro ser asp tyr asp arg leu pro val ser gly ala val val leu val
601/201                               631/211
gat cgc ggc gtc tgt cct ttt gcc cag aag gaa gac gca gcc gcg cag cgc ggt gcg gtg
asp arg gly val cys pro phe ala gln lys glu asp ala ala ala gln arg gly ala val
661/221                               691/231
gcg ctg atc att gct gac aac atc gac gag cag gcg atg ggc ggc acc ctg ggg gct aat
ala leu ile ile ala asp asn ile asp glu gln ala met gly gly thr leu gly ala asn
721/241                               751/251
acc gac gtc aag atc ccg gtg gtg agt gtc acc aag tcg gtc gga ttc cag cta cgc gga
thr asp val lys ile pro val val ser val thr lys ser val gly phe gln leu arg gly
781/261                               811/271
cag tct ggg cca acc acc gtc aag ctc acg gcg agc acc caa agt ttc aag gcc cgc aac
gln ser gly pro thr thr val lys leu thr ala ser thr gln ser phe lys ala arg asn
841/281                               871/291
gtc atc gcg cag acg aag acg ggg tcg tcg gcc aac gtg gtg atg gca ggt gcg cat ttg
val ile ala gln thr lys thr gly ser ser ala asn val val met ala gly ala his leu
901/301                               931/311
gac agc gtt ccg gaa gga ccc ggc atc aac gac aac ggc tcg gga gtg gct gcg gtt ctg
asp ser val pro glu gly pro gly ile asn asp asn gly ser gly val ala ala val leu
961/321                               991/331
gaa acg gca gtg cag ctg ggg aac tca ccg cat gtg tcc aac gcg gta cgg ttc gcc ttc
glu thr ala val gln leu gly asn ser pro his val ser asn ala val arg phe ala phe

```

SEQ ID N° 19 F

FIGURE 19F
FEUILLE DE REMPLACEMENT (REGLE 26)

73/185

1021/341
 tgg ggc gcc gag gaa ttc ggc ctg att ggg tca cga aac tac gtc gag tcg ctg gac atc
 trp gly ala glu glu phe gly leu ile gly ser arg asn tyr val glu ser leu asp ile
 1081/361
 gac gcg ctc aaa ggc atc gcg ctg tat ctg aac ttc gac atg ttg gcg tcg ccg aac ccg
 asp ala leu lys gly ile ala leu tyr leu asn phe asp met leu ala ser pro asn pro
 1141/381
 ggt tac ttc acc tac gac ggt gac cag tcg ctg ccg cta gac gcc cgc ggt cag ccg gtg
 gly tyr phe thr tyr asp gly asp gln ser leu pro leu asp ala arg gly gln pro val
 1201/401
 gtg ccc gaa ggc tcg gcc ggt atc gag cgc acg ttc gtc gcc tat ctg aag atg gcc gcc
 val pro glu gly ser ala gly ile glu arg thr phe val ala tyr leu lys met ala gly
 1261/421
 aag acc gcg cag gac acc tcg ttc gac ggt cgg tcc gac tac gac ggc ttc acg ctg gcg
 lys thr ala gln asp thr ser phe asp gly arg ser asp tyr asp gly phe thr leu ala
 1321/441
 ggt atc cct tcg ggt ggc ctg ttc tcc gcc gct gag gtc aag aag tcc gcc gag caa gcc
 gly ile pro ser gly gly leu phe ser gly ala glu val lys lys ser ala glu gln ala
 1381/461
 gag ctc tgg ggc gcc acc gcc gac gag cct ttc gat ccc aac tat cac cag aag aca gac
 glu leu trp gly gly thr ala asp glu pro phe asp pro asn tyr his gln lys thr asp
 1441/481
 acc ctg gac cat atc gac cgc acc gcg ctc ggt atc aac ggc gct ggc gtc gcg tac gcg
 thr leu asp his ile asp arg thr ala leu gly ile asn gly ala gly val ala tyr ala
 1501/501
 gtg ggt ttg tat gcg cag gac ctc ggc gcc ccc aac ggg gtt ccg gtc atg gcg gac cgc
 val gly leu tyr ala gln asp leu gly gly pro asn gly val pro val met ala asp arg
 1561/521
 acc cgc cac ctg att gcc aaa ccg tga
 thr arg his leu ile ala lys pro OPA

SEQ ID N° 19F (suite)

FIGURE 19F (suite)

31/11
 CGA GAC AGT GGT GCG GGA CAC TTG AGT TCG GCT GCT AAC GAC GCC AGA GTC GCC CGC TTC
 arg asp ser gly ala gly his leu ser ser ala ala asn asp ala arg val ala arg phe
 61/21
 CGC GGT GTG GGA CTC ACG TTC GGT GAG GGT ACA CGC GAC CTT CGA GCA CGC AAT ATC GTG
 arg gly val gly leu thr phe gly glu gly thr ala asp leu arg ala arg asn ile val
 121/41
 GGC CGG CTG GCA ACC GTC GGT TTC GAC GTT GGT GAC GAC CCC TCG TTC ATG AAT CGT TCT
 gly arg leu ala thr val gly phe asp val gly asp asp pro ser phe met asn arg ser
 181/61
 TGA GCT CCC CGT TTT GCT GGA TGC CCA GGC ACC GCC GGT ACT GCT GCG CTT AAG CTT GTC
 OPA ala pro arg phe ala gly cys pro gly thr ala gly thr ala ala leu lys leu val
 241/81
 GCA CAT GGT GCC GGC AGG GAG GAA CAG TGG GCA AGC AGC TAG CCG CGC TCG CCG CGC TGG
 ala his gly ala gly arg glu glu gln trp ala ser ser AMB pro arg ser pro arg trp
 301/101
 TCG GTG CGT GCA TGC TCG CAG CCG GAT GCA CCA ACG TGG TCG ACG GGA CCG CCG TGG CTG
 ser val arg ala cys ser gln pro asp ala pro thr trp ser thr gly pro pro trp leu
 361/121
 CCG ACA AAT CCG GAC CAC TGC ATC AGG ATC
 pro thr asn pro asp his cys ile arg ile

SEQ ID N° 20A

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 20A

74/185

32/11

GAG ACA GTG GTG CGG GAC ACT TGA GTT CGG CTG CTA ACG ACG CCA GAG TCG CCC GCT TCC
 glu thr val val arg asp thr OPA val arg leu leu thr thr pro glu ser pro ala ser
 62/21 92/31

GCG GTG TGG GAC TCA CGT TCG GTG AGG GTA CAG CGG ACC TTC GAG CAC GCA ATA TCG TGG
 ala val trp asp ser arg ser val arg val gln arg thr phe glu his ala ile ser trp
 122/41 152/51

GCC GGC TGG CAA CCG TCG GTT TCG ACG TTG GTG ACG ACC CCT CGT TCA TGA ATC GTT CTT
 ala gly trp gln pro ser val ser thr leu val thr thr pro arg ser OPA ile val leu
 182/61 212/71

GAG CTC CCC GTT TTG CTG GAT GCC CAG GCA CCG CCG GTA CTG CTG CGC TTA AGC TTG TCG
 glu leu pro val leu leu asp ala gln ala pro pro val leu leu arg leu ser leu ser
 242/81 272/91

CAC ATG GTG CCG GCA GGG AGG AAC AGT GGG CAA GCA GCT AGC CGC GCT CGC CGC GCT GGT
 his met val pro ala gly arg asn ser gly gln ala ala ser arg ala arg arg ala gly
 302/101 332/111

CGG TGC GTG CAT GCT CGC AGC CGG ATG CAC CAA CGT GGT CGA CGG GAC CGC CGT GGC TGC
 arg cys val his ala arg ser arg met his gln arg gly arg arg asp arg arg gly cys
 362/121

CGA CAA ATC CGG ACC ACT GCA TCA GGA TC
 arg gln ile arg thr thr ala ser gly

SEQ ID N° 20B

FIGURE 20B

33/11

AGA CAG TGG TGC GGG ACA CTT GAG TTC GGC TGC TAA CGA CGC CAG AGT CGC CCG CTT CCG
 arg gln trp cys gly thr leu glu phe gly cys OCH arg arg gln ser arg pro leu pro
 63/21 93/31

CGG TGT GGG ACT CAC GTT CGG TGA GGG TAC AGC GGA CCT TCG AGC ACG CAA TAT CGT GGG
 arg cys gly thr his val arg OPA gly tyr ser gly pro ser ser thr gln tyr arg gly
 123/41 153/51

CCG GCT GGC AAC CGT CGG TTT CGA CGT TGG TGA CGA CCC CTC GTT CAT GAA TCG TTC TTG
 pro ala gly asn arg arg phe arg arg trp OPA arg pro leu val his glu ser phe leu
 183/61 213/71

AGC TCC CCG TTT TGC TGG ATG CCC AGG CAC CGC CGG TAC TGC TGC GCT TAA GCT TGT CGC
 ser ser pro phe cys trp met pro arg his arg arg tyr cys cys ala OCH ala cys arg
 243/81 273/91

ACA TGG TGC CGG CAG GGA GGA ACA GTG GGC AAG CAG CTA GCC GCG CTC GCC GCG CTG GTC
 thr trp cys arg gln gly gly thr val gly lys gln leu ala ala leu ala ala leu val
 303/101 333/111

GGT GCG TGC ATG CTC GCA GCC GGA TGC ACC AAC GTG GTC GAC GGG ACC GCC GTG GCT GCC
 gly ala cys met leu ala ala gly cys thr asn val val asp gly thr ala val ala ala
 363/121

GAC AAA TCC GGA CCA CTG CAT CAG GAT C
 asp lys ser gly pro leu his gln asp

SEQ ID N° 20C

FIGURE 20C
 FEUILLE DE REMPLACEMENT (REGLE 26)

75/185

partie de la séquence nucléotidique de seq20A

```

1/1                               31/11
TGT GGG ACT CAC GTT CGG TGA GGG TAC AGC GGA CCT TCG AGC ACG CAA TAT CGT GGG CCG
cys gly thr his val arg OPA gly tyr ser gly pro ser ser thr gln tyr arg gly pro
61/21                               91/31
GCT GGC AAC CGT CGG TTT CGA CGT TGG TGA CGA CCC CTC GTT CAT GAA TCG TTC TTG AGC
ala gly asn arg arg phe arg arg trp OPA arg pro leu val his glu ser phe leu ser
121/41                               151/51
TCC CCG TTT TGC TGG ATG CCC AGG CAC CGC CGG TAC TGC TGC GCT TAA GCT TGT CGC ACA
ser pro phe cys trp met pro arg his arg arg tyr cys cys ala OCH ala cys arg thr
181/61                               211/71
TGG TGC CGG CAG GGA GGA ACA GTG GGC AAG CAG CTA GCC GCG CTC GCC GCG CTG GTC GGT
trp cys arg gln gly gly thr val gly lys gln leu ala ala leu ala ala leu val gly
241/81                               271/91
GCG TGC ATG CTC GCA GCC GGA TGC ACC AAC GTG GTC GAC GGG ACC GCC GTG GCT GCC GAC
ala cys met leu ala ala gly cys thr asn val val asp gly thr ala val ala ala asp
301/101
AAA TCC GGA CCA CTG CAT CAG GAT C
lys ser gly pro leu his gln asp

```

SEQ ID N° 20A'

FIGURE 20A'

```

1/1                               31/11
GTG GGA CTC ACG TTC GGT GAG GGT ACA GCG GAC CTT CGA GCA CGC AAT ATC GTG GGC CGG
val gly leu thr phe gly glu gly thr ala asp leu arg ala arg asn ile val gly arg
61/21                               91/31
CTG GCA ACC GTC GGT TTC GAC GTT GGT GAC GAC CCC TCG TTC ATG AAT CGT TCT TGA GCT
leu ala thr val gly phe asp val gly asp asp pro ser phe met asn arg ser OPA ala
121/41                               151/51
CCC CGT TTT GCT GGA TGC CCA GGC ACC GCC GGT ACT GCT GCG CTT AAG CTT GTC GCA CAT
pro arg phe ala gly cys pro gly thr ala gly thr ala ala leu lys leu val ala his
181/61                               211/71
GGT GCC GGC AGG GAG GAA CAG TGG GCA AGC AGC TAG CCG CGC TCG CCG CGC TGG TCG GTG
gly ala gly arg glu glu gln trp ala ser ser AMB pro arg ser pro arg trp ser val
241/81                               271/91
CGT GCA TGC TCG CAG CCG GAT GCA CCA ACG TGG TCG ACG GGA CCG CCG TGG CTG CCG ACA
arg ala cys ser gln pro asp ala pro thr trp ser thr gly pro pro trp leu pro thr
301/101
AAT CCG GAC CAC TGC ATC AGG ATC
asn pro asp his cys ile arg ile

```

SEQ ID N° 20B'

FIGURE 20B'

76/185

1/1 31/11
 GTG TGG GAC TCA CGT TCG GTG AGG GTA CAG CGG ACC TTC GAG CAC GCA ATA TCG TGG GCC
 val trp asp ser arg ser val arg val gln arg thr phe glu his ala ile ser trp ala
 61/21 91/31
 GGC TGG CAA CCG TCG GTT TCG ACG TTG GTG ACG ACC CCT CGT TCA TGA ATC GTT CTT GAG
 gly trp gln pro ser val ser thr leu val thr thr pro arg ser OPA ile val leu glu
 121/41 151/51
 CTC CCC GTT TTG CTG GAT GCC CAG GCA CCG CCG GTA CTG CTG CGC TTA AGC TTG TCG CAC
 leu pro val leu leu asp ala gln ala pro pro val leu leu arg leu ser leu ser his
 181/61 211/71
 ATG GTG CCG GCA GGG AGG AAC AGT GGG CAA GCA GCT AGC CGC GCT CGC CGC GCT GGT CGG
 met val pro ala gly arg asn ser gly gln ala ala ser arg ala arg arg ala gly arg
 241/81 271/91
 TGC GTG CAT GCT CGC AGC CGG ATG CAC CAA CGT GGT CGA CGG GAC CGC CGT GGC TGC CGA
 cys val his ala arg ser arg met his gln arg gly arg arg asp arg arg gly cys arg
 301/101
 CAA ATC CGG ACC ACT GCA TCA GGA TC
 gln ile arg thr thr ala ser gly

SEQ ID N° 20C'

FIGURE 20C'

séquence Rv3576 prédite par Cole et al. (Nature 393:537-544) et contenant seq20A'
 1/1 31/11
 atg ggc aag cag cta gcc gcg ctc gcc gcg ctg gtc ggt gcg tgc atg ctc gca gcc gga
 met gly lys gln leu ala ala leu ala ala leu val gly ala cys met leu ala ala gly
 61/21 91/31
 tgc acc aac gtg gtc gac ggg acc gcc gtg gct gcc gac aaa tcc gga cca ctg cat cag
 cys thr asn val val asp gly thr ala val ala ala asp lys ser gly pro leu his gln
 121/41 151/51
 gat ccg ata ccg gtt tca gcg ctt gaa ggg ctg ctt ctc gac ttg agc cag atc aat gcc
 asp pro ile pro val ser ala leu glu gly leu leu leu asp leu ser gln ile asn ala
 181/61 211/71
 gcg ctg ggt gcg aca tcg atg aag gtg tgg ttc aac gcc aag gca atg tgg gac tgg agc
 ala leu gly ala thr ser met lys val trp phe asn ala lys ala met trp asp trp ser
 241/81 271/91
 aag agc gtg gcc gac aag aat tgc ctg gct atc gac ggt cca gca cag gaa aag gtc tat
 lys ser val ala asp lys asn cys leu ala ile asp gly pro ala gln glu lys val tyr
 301/101 331/111
 gcc ggc acc ggg tgg acc gct atg cgc gcc caa ccg ctg gat gac agc atc gat gac tcc
 ala gly thr gly trp thr ala met arg gly gln arg leu asp asp ser ile asp asp ser
 361/121 391/131
 aag aaa cgc gac cac tac gcc att caa gcg gtc gtc ggc ttc ccg acc gca cat gat gcc
 lys lys arg asp his tyr ala ile gln ala val val gly phe pro thr ala his asp ala
 421/141 451/151
 gag gag ttc tac agc tcc tcg gtg caa agc tgg agc agc tgc tcg aac cgc ccg ttt gtc
 glu glu phe tyr ser ser ser val gln ser trp ser ser cys ser asn arg arg phe val
 481/161 511/171
 gaa gtc acc ccc gga cag gac gac gcc gcc tgg act gtg gct gac gtt gtc aac gac aac
 glu val thr pro gly gln asp asp ala ala trp thr val ala asp val val asn asp as
 541/181 571/191
 ggc atg ctc agt agc tcg cag gtt cag gaa ggc ggc gac gga tgg acc tgc cag cgt gcc
 gly met leu ser ser ser gln val gln glu gly gly asp gly trp thr cys gln arg ala
 601/201 631/211
 ctg act gcg cgc aac aac gtc act atc gac att gtc acg tgc gcc tat agc caa ccg gat
 leu thr ala arg asn asn val thr ile asp ile val thr cys ala tyr ser gln pro asp
 661/221 691/231
 ttg gtg gcg att ggc atc gct aac caa atc gcg gcc aag gtt gct aag cag tag
 leu val ala ile gly ile ala asn gln ile ala ala lys val ala lys gln AMB

SEQ ID N° 20D

FIGURE 20D
 FEUILLE DE REMPLACEMENT (REGLE 26)

77/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv3576

```

1/1                                31/11
taa gct tgt cgc aca tgg tgc cgg cag gga gga aca gtg ggc aag cag cta gcc gcg ctc
OCH ala cys arg thr trp cys arg gln gly gly thr val gly lys gln leu ala ala leu
61/21                                91/31
gcc gcg ctg gtc ggt gcg tgc atg ctc gca gcc gga tgc acc aac gtg gtc gac ggg acc
ala ala leu val gly ala cys met leu ala ala gly cys thr asn val val asp gly thr
121/41                                151/51
gcc gtg gct gcc gac aaa tcc gga cca ctg cat cag gat ccg ata ccg gtt tca gcg ctt
ala val ala ala asp lys ser gly pro leu his gln asp pro ile pro val ser ala leu
181/61                                211/71
gaa ggg ctg ctt ctc gac ttg agc cag atc aat gcc gcg ctg ggt gcg aca tcg atg aag
glu gly leu leu leu asp leu ser gln ile asn ala ala leu gly ala thr ser met lys
241/81                                271/91
gtg tgg ttc aac gcc aag gca atg tgg gac tgg agc aag agc gtg gcc gac aag aat tgc
val trp phe asn ala lys ala met trp asp trp ser lys ser val ala asp lys asn cys
301/101                               331/111
ctg gct atc gac ggt cca gca cag gaa aag gtc tat gcc ggc acc ggg tgg acc gct atg
leu ala ile asp gly pro ala gln glu lys val tyr ala gly thr gly trp thr ala met
361/121                               391/131
cgc ggc caa cgg ctg gat gac agc atc gat gac tcc aag aaa cgc gac cac tac gcc att
arg gly gln arg leu asp asp ser ile asp asp ser lys lys arg asp his tyr ala ile
421/141                               451/151
caa gcg gtc gtc ggc ttc ccg acc gca cat gat gcc gag gag ttc tac agc tcc tcg gtg
gln ala val val gly phe pro thr ala his asp ala glu glu phe tyr ser ser ser val
481/161                               511/171
caa agc tgg agc agc tgc tcg aac cgc cgg ttt gtc gaa gtc acc ccc gga cag gac gac
gln ser trp ser ser cys ser asn arg arg phe val glu val thr pro gly gln asp asp
541/181                               571/191
gcc gcc tgg act gtg gct gac gtt gtc aac gac aac ggc atg ctc agt agc tcg cag gtt
ala ala trp thr val ala asp val val asn asp asn gly met leu ser ser ser gln val
601/201                               631/211
cag gaa ggc ggc gac gga tgg acc tgc cag cgt gcc ctg act gcg cgc aac aac gtc act
gln glu gly gly asp gly trp thr cys gln arg ala leu thr ala arg asn asn val thr
661/221                               691/231
atc gac att gtc acg tgc gcc tat agc caa ccg gat ttg gtg gcg att ggc atc gct aac
ile asp ile val thr cys ala tyr ser gln pro asp leu val ala ile gly ile ala asn
721/241
caa atc gcg gcc aag gtt gct aag cag tag
gln ile ala ala lys val ala lys gln AMB

```

SEQ ID N° 20F

FIGURE 20F

78/185

1/1 31/11
 GTC CTG GTC GCC GCG CAA CTG GCC GGT CCC GAT GGA AAG TGT TCA CGA TCG CGC TTC TGC
 val leu val ala ala gln leu ala gly pro asp gly lys cys ser arg ser arg phe cys
 61/21 * 91/31
 CGC TGG TAG TGG CGA TGG TGT TAG CAG GAT TGC GGG TCG AGG CTG CGA TGG CCA GCA CCA
 arg trp AMB trp arg trp cys AMB gln asp cys gly ser arg leu arg trp pro ala pro
 121/41 151/51
 GCG GCC TGC GGC TGG TCG CCG CGC GCG CCG AAA TGA TAC CCG CGA TCA CGA AAT ACA TGT
 ala ala cys gly trp ser pro arg ala pro lys OPA tyr pro arg ser arg asn thr cys
 181/61 211/71
 CGG CGC TGG ACG TCG CCG TGC TGG CCA GCT CGA CCG GAC ACG ATG TGG AGG GGG CGC AGA
 arg arg trp thr ser pro cys trp pro ala arg pro asp thr met trp arg gly arg arg
 241/81 271/91
 AAA ACT TCA CCG CCC GCA AGT ACG AGC TGC AGA CGC GAC TGG CCG ACA CCG ACG TCA TCG
 lys thr ser pro pro ala ser thr ser cys arg arg asp trp pro thr pro thr ser ser
 301/101 331/111
 CAG ACG TGC GGT CGG GAG TGA ACA CGC TGC TCA ACG GCG GTC AGG CGC TGC TGG ATA AGA
 gln thr cys gly arg glu OPA thr arg cys ser thr ala val arg arg cys trp ile arg
 361/121
 TGC TGG CCG ACA GCA TCG GCT TGC GGG ATC
 cys trp pro thr ala ser ala cys gly ile

SEQ ID N° 21A

FIGURE 21A

32/11
 TCC TGG TCG CCG CGC AAC TGG CCG GTC CCG ATG GAA AGT GTT CAC GAT CGC GCT TCT GCC
 ser trp ser pro arg asn trp pro val pro met glu ser val his asp arg ala ser ala
 62/21 92/31
 GCT GGT AGT GGC GAT GGT GTT AGC AGG ATT GCG GGT CGA GGC TGC GAT GGC CAG CAC CAG
 ala gly ser gly asp gly val ser arg ile ala gly arg gly cys asp gly gln his gln
 122/41 152/51
 CGG CCT GCG GCT GGT CGC CGC GCG CGC CGA AAT GAT ACC CGC GAT CAC GAA ATA CAT GTC
 arg pro ala ala gly arg arg ala arg arg asn asp thr arg asp his glu ile his val
 182/61 212/71
 GGC GCT GGA CGT CGC CGT GCT GGC CAG CTC GAC CGG ACA CGA TGT GGA GGG GGC GCA GAA
 gly ala gly arg arg arg ala gly gln leu asp arg thr arg cys gly gly gly ala glu
 242/81 272/91
 AAA CTT CAC CGC CCG CAA GTA CGA GCT GCA GAC GCG ACT GGC CGA CAC CGA CGT CAT CGC
 lys leu his arg pro gln val arg ala ala asp ala thr gly arg his arg arg his arg
 302/101 332/111
 AGA CGT GCG GTC GGG AGT GAA CAC GCT GCT CAA CGG CGG TCA GGC GCT GCT GGA TAA GAT
 arg arg ala val gly ser glu his ala ala gln arg arg ser gly ala ala gly OCH asp
 362/121
 GCT GGC CGA CAG CAT CGG CTT GCG GGA TC
 ala gly arg gln his arg leu ala gly

SEQ ID N° 21B

FIGURE 21B

79/185

33/11
 CCT GGT CGC CGC GCA ACT GGC CGG TCC CGA TGG AAA GTG TTC ACG ATC GCG CTT CTG CCG
 pro gly arg arg ala thr gly arg ser arg trp lys val phe thr ile ala leu leu pro
 63/21 93/31
 CTG GTA GTG GCG ATG GTG TTA GCA GGA TTG CGG GTC GAG GCT GCG ATG GCC AGC ACC AGC
 leu val val ala met val leu ala gly leu arg val glu ala ala met ala ser thr ser
 123/41 153/51
 GGC CTG CGG CTG GTC GCC GCG CGC GCC GAA ATG ATA CCC GCG ATC ACG AAA TAC ATG TCG
 gly leu arg leu val ala ala arg ala glu met ile pro ala ile thr lys tyr met ser
 183/61 213/71
 GCG CTG GAC GTC GCC GTG CTG GCC AGC TCG ACC GGA CAC GAT GTG GAG GGG GCG CAG AAA
 ala leu asp val ala val leu ala ser ser thr gly his asp val glu gly ala gln lys
 243/81 273/91
 AAC TTC ACC GCC CGC AAG TAC GAG CTG CAG ACG CGA CTG GCC GAC ACC GAC GTC ATC GCA
 asn phe thr ala arg lys tyr glu leu gln thr arg leu ala asp thr asp val ile ala
 303/101 333/111
 GAC GTG CGG TCG GGA GTG AAC ACG CTG CTC AAC GGC GGT CAG GCG CTG CTG GAT AAG ATG
 asp val arg ser gly val asn thr leu leu asn gly gly gln ala leu leu asp lys met
 363/121
 CTG GCC GAC AGC ATC GGC TTG CGG GAT C
 leu ala asp ser ile gly leu arg asp

SEQ ID N° 21C

FIGURE 21C

partie de la séquence nucléotidique de seq21A

1/1 31/11
 ACG ATC GCG CTT CTG CCG CTG GTA GTG GCG ATG GTG TTA GCA GGA TTG CGG GTC GAG GCT
 thr ile ala leu leu pro leu val val ala met val leu ala gly leu arg val glu ala
 61/21 91/31
 GCG ATG GCC AGC ACC AGC GGC CTG CGG CTG GTC GCC GCG CGC GCC GAA ATG ATA CCC GCG
 ala met ala ser thr ser gly leu arg leu val ala ala arg ala glu met ile pro ala
 121/41 151/51
 ATC ACG AAA TAC ATG TCG GCG CTG GAC GTC GCC GTG CTG GCC AGC TCG ACC GGA CAC GAT
 ile thr lys tyr met ser ala leu asp val ala val leu ala ser ser thr gly his asp
 181/61 211/71
 GTG GAG GGG GCG CAG AAA AAC TTC ACC GCC CGC AAG TAC GAG CTG CAG ACG CGA CTG GCC
 val glu gly ala gln lys asn phe thr ala arg lys tyr glu leu gln thr arg leu ala
 241/81 271/91
 GAC ACC GAC GTC ATC GCA GAC GTG CGG TCG GGA GTG AAC ACG CTG CTC AAC GGC GGT CAG
 asp thr asp val ile ala asp val arg ser gly val asn thr leu leu asn gly gly gln
 301/101 331/111
 GCG CTG CTG GAT AAG ATG CTG GCC GAC AGC ATC GGC TTG CGG GAT C
 ala leu leu asp lys met leu ala asp ser ile gly leu arg asp

SEQ ID N° 21A'

FIGURE 21A'

80/185

1/1 31/11
 CGA TCG CGC TTC TGC CGC TGG TAG TGG CGA TGG TGT TAG CAG GAT TGC GGG TCG AGG CTG
 arg ser arg phe cys arg trp AMB trp arg trp cys AMB gln asp cys gly ser arg leu
 61/21 91/31
 CGA TGG CCA GCA CCA GCG GCC TGC GGC TGG TCG CCG CGC GCG CCG AAA TGA TAC CCG CGA
 arg trp pro ala pro ala ala cys gly trp ser pro arg ala pro lys OPA tyr pro arg
 121/41 151/51
 TCA CGA AAT ACA TGT CGG CGC TGG ACG TCG CCG TGC TGG CCA GCT CGA CCG GAC ACG ATG
 ser arg asn thr cys arg arg trp thr ser pro cys trp pro ala arg pro asp thr met
 181/61 211/71
 TGG AGG GGG CGC AGA AAA ACT TCA CCG CCC GCA AGT ACG AGC TGC AGA CGC GAC TGG CCG
 trp arg gly arg arg lys thr ser pro pro ala ser thr ser cys arg arg asp trp pro
 241/81 271/91
 ACA CCG ACG TCA TCG CAG ACG TGC GGT CGG GAG TGA ACA CGC TGC TCA ACG GCG GTC AGG
 thr pro thr ser ser gln thr cys gly arg glu OPA thr arg cys ser thr ala val arg
 301/101 331/111
 CGC TGC TGG ATA AGA TGC TGG CCG ACA GCA TCG GCT TGC GGG ATC
 arg cys trp ile arg cys trp pro thr ala ser ala cys gly ile

SEQ ID N° 21B'

FIGURE 21B'

1/1 31/11
 CAC GAT CGC GCT TCT GCC GCT GGT AGT GGC GAT GGT GTT AGC AGG ATT GCG GGT CGA GGC
 his asp arg ala ser ala ala gly ser gly asp gly val ser arg ile ala gly arg gly
 61/21 91/31
 TGC GAT GGC CAG CAC CAG CGG CCT GCG GCT GGT CGC CGC GCG CGC CGA AAT GAT ACC CGC
 cys asp gly gln his gln arg pro ala ala gly arg arg ala arg arg asn asp thr arg
 121/41 151/51
 GAT CAC GAA ATA CAT GTC GGC GCT GGA CGT CGC CGT GCT GGC CAG CTC GAC CGG ACA CGA
 asp his glu ile his val gly ala gly arg arg arg ala gly gln leu asp arg thr arg
 181/61 211/71
 TGT GGA GGG GGC GCA GAA AAA CTT CAC CGC CCG CAA GTA CGA GCT GCA GAC GCG ACT GGC
 cys gly gly gly ala glu lys leu his arg pro gln val arg ala ala asp ala thr gly
 241/81 271/91
 CGA CAC CGA CGT CAT CGC AGA CGT GCG GTC GGG AGT GAA CAC GCT GCT CAA CGG CGG TCA
 arg his arg arg his arg arg arg ala val gly ser glu his ala ala gln arg arg ser
 301/101 331/111
 GGC GCT GCT GGA TAA GAT GCT GGC CGA CAG CAT CGG CTT GCG GGA TC
 gly ala ala gly OCH asp ala gly arg gln his arg leu ala gly

SEQ ID N° 21C'

FIGURE 21C'

81/185

séquence Rv3365c prédite par Cole et al. (Nature 393:537-544) et contenant Seq21A'

1/1	31/11
gtg acc atg ttc gcc cgc ccg acc atc ccg	gtc gcg gcg gcc gct tct gat att tcc gcc
val thr met phe ala arg pro thr ile pro	val ala ala ala ala ser asp ile ser ala
61/21	91/31
ccg gct caa ccg gcc cgc ggc aaa cct cag	caa cgc ccg ccg tcc tgg tcg ccg cgc aac
pro ala gln pro ala arg gly lys pro gln	gln arg pro pro ser trp ser pro arg asn
121/41	151/51
tgg ccg gtc cga tgg aaa gtg ttc acg atc	gcg ctt ctg ccg ctg gta gtg gcg atg gtg
trp pro val arg trp lys val phe thr ile	ala leu leu pro leu val val ala met val
181/61	211/71
tta gca gga ttg cgg gtc gag gct gcg atg	gcc agc acc agc ggc ctg cgg ctg gtc gcc
leu ala gly leu arg val glu ala ala met	ala ser thr ser gly leu arg leu val ala
241/81	271/91
gcg cgc gcc gaa atg ata ccc gcg atc acg	aaa tac atg tcg gcg ctg gac gtc gcc gtg
ala arg ala glu met ile pro ala ile thr	lys tyr met ser ala leu asp val ala val
301/101	331/111
ctg gcc agc tcg acc gga cac gat gtg gag	ggg gcg cag aaa aac ttc acc gcc cgc aag
leu ala ser ser thr gly his asp val glu	gly ala gln lys asn phe thr ala arg lys
361/121	391/131
tac gag ctg cag acg cga ctg gcc gac acc	gac gtc atc gca gac gtg cgg tcg gga gtg
tyr glu leu gln thr arg leu ala asp thr	asp val ile ala asp val arg ser gly val
421/141	451/151
aac acg ctg ctc aac ggc ggt cag gcg ctg	ctg gat aag gtg ctg gcc gac agc atc ggc
asn thr leu leu asn gly gly gln ala leu	leu asp lys val leu ala asp ser ile gly
481/161	511/171
ttg ccg gat ccg gtc acc gcc tac gcg ccg	ctg ctg ttg acg gcc cag aac gtg att gac
leu arg asp arg val thr ala tyr ala pro	leu leu leu thr ala gln asn val ile asp
541/181	571/191
gcg tcg gtg ccg gtt gac agc gag caa atc	cga acc cag gtg cag ggt ttg agc cga gcc
ala ser val arg val asp ser glu gln ile	arg thr gln val gln gly leu ser arg ala
601/201	631/211
ggt ggc gcc cgc ggg cag atg acg atg cag	gag atc ctg gtg act cgc ggc gcc gac ctt
val gly ala arg gly gln met thr met gln	glu ile leu val thr arg gly ala asp leu
661/221	691/231
gcc gag ccg caa ctg cgc agc gcg atg gtt	acc ctg gcc ggc acc gaa ccc tcg acg ctg
ala glu pro gln leu arg ser ala met val	thr leu ala gly thr glu pro ser thr leu
721/241	751/251
ttc ggg atg agc gcg gcg ctc ggt gca ggc	tcg ccg gac acc aag aac ctg cag cag caa
phe gly met ser ala ala leu gly ala gly	ser pro asp thr lys asn leu gln gln gln
781/261	811/271
atg gtg acc agg atg gcg atc atg tcc gat	ccg gcc gtt gca ctg gtc aac aac cca gag
met val thr arg met ala ile met ser asp	pro ala val ala leu val asn asn pro glu
841/281	871/291
ctg ctg cac tcg ata cag atc acc cgc gac	att gcc gag cag gtg atc acc gac acc acc
leu leu his ser ile gln ile thr arg asp	ile ala glu gln val ile thr asp thr thr
901/301	931/311
gag gcg gtg acg aag tcg gtg caa agc cag	gcc acc gac ccg ccg gat gcc gcg att cgc
glu ala val thr lys ser val gln ser gln	ala thr asp arg arg asp ala ala ile arg
961/321	991/331
gac gcc gtg ctg gtg ttg gcc gcc atc gcg	acc gcg atc gtc gtc gtg ttg gtg gtg gcg
asp ala val leu val leu ala ala ile ala	thr ala ile val val val leu val val ala

SEQ ID N° 21F

FIGURE 21D
FEUILLE DE REMPLACEMENT (REGLE 26)

82/185

1021/341	1051/351
cgc acg ctg gtc ggg ccg atg cgg gta ctg	cgt gat ggg gcg ctc aag gtt gct cat acc
arg thr leu val gly pro met arg val leu	arg asp gly ala leu lys val ala his thr
1081/361	1111/371
gat ctc gac ggc gag atc gcg gcg gtc cgc	gcc ggc gac gag ccg atc ccc gag cca ctg
asp leu asp gly glu ile ala ala val arg	ala gly asp glu pro ile pro glu pro leu
1141/381	1171/391
gcg gtg tac acc acc gag gaa atc ggt cag	gtc gcg cat gcg gtc gac gag ctg cac acc
ala val tyr thr thr glu glu ile gly gln	val ala his ala val asp glu leu his thr
1201/401	1231/411
cgg gcc ctg ttg ctg gcc ggc gag gaa acg	cgg ttg cga ctg ctg gtc aac gag atg ttt
arg ala leu leu leu ala gly glu glu thr	arg leu arg leu leu val asn glu met phe
1261/421	1291/431
gag acc atg tcg cgg cgt agc cgt tcc ctg	gtc gac cag cag ctg tcg gtc atc gac caa
glu thr met ser arg arg ser arg ser leu	val asp gln gln leu ser val ile asp gln
1321/441	1351/451
ctg gag cgc aac gag gag gat ccc gcc cga	ctc gac agc ctt ttc cgg ctc gat cac ctg
leu glu arg asn glu glu asp pro ala arg	leu asp ser leu phe arg leu asp his leu
1381/461	1411/471
gcc gcc cgg ctg cgc cgc aac agc gcc aac	ctg ctg gtg ctg gcc ggt gcg cag att acc
ala ala arg leu arg arg asn ser ala asn	leu leu val leu ala gly ala gln ile thr
1441/481	1471/491
cgt gac cac cgc gag ccg gtg ccg ctg tca	acc gtg atc agc gcc gcc gtg tca gag gtc
arg asp his arg glu pro val pro leu ser	thr val ile ser ala ala val ser glu val
1501/501	1531/511
gag gac tat cgc cgc gtc gac atc gcg agg	gta ccc gac tgt gcg gta gtc ggc gca gcg
glu asp tyr arg arg val asp ile ala arg	val pro asp cys ala val val gly ala ala
1561/521	1591/531
gct ggt ggc gtc att cat ctg ctt gcc gag	ctg atc gac aac gcg ttg cgc tac tcg tca
ala gly gly val ile his leu leu ala glu	leu ile asp asn ala leu arg tyr ser ser
1621/541	1651/551
ccg acc aca ccc gtt cgg gtt gcc gcc gca	atc ggc agc gaa ggc agt gtt ctg ctg cga
pro thr thr pro val arg val ala ala ala	ile gly ser glu gly ser val leu leu arg
1681/561	1711/571
atc tcg gat tcc ggc ctg ggc atg acc gat	gcc gat cgg cgg atg gcc aat atg cgg ctg
ile ser asp ser gly leu gly met thr asp	ala asp arg arg met ala asn met arg leu
1741/581	1771/591
cgg gcc ggc ggt gag gtc acc ccg gat agt	gcc cgg cac atg ggt ctg ttc gta gtc gcc
arg ala gly gly glu val thr pro asp ser	ala arg his met gly leu phe val val gly
1801/601	1831/611
cgg ctg gcc ggt cgg cac ggc atc cga gtc	ggg ctg cgc ggt ccg gtg acc ggt gaa cag
arg leu ala gly arg his gly ile arg val	gly leu arg gly pro val thr gly glu gln
1861/621	1891/631
ggc acc ggc acc acc gcc gag gtc tac ctg	ccg cta gcc gtg ctc gag ggg acg gcc cca
gly thr gly thr thr ala glu val tyr leu	pro leu ala val leu glu gly thr ala pro
1921/641	1951/651
gcg cag ccg cca aag ccg cgg gta ttt gcg	atc aag ccg ccg tgt cct gaa ccc gcg gcg
ala gln pro pro lys pro arg val phe ala	ile lys pro pro cys pro glu pro ala ala
1981/661	2011/671
gcc gat ccg acg gac gtt ccc gcc gcc atc	ggg ccg cta cca ccg gtc acg ttg ctc ccg
ala asp pro thr asp val pro ala ala ile	gly pro leu pro pro val thr leu leu pro

SEQ ID N° 21D (suite 1)

FIGURE 21D (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

83/185

2041/681	2071/691
cgc cgt acc ccg ggg tcc agt ggc atc gcc	gac gtc ccg gcc cag ccg atg cag cag ccg
arg arg thr pro gly ser ser gly ile ala	asp val pro ala gln pro met gln gln arg
2101/701	2131/711
cgc cgc gag ctg aaa aca ccc tgg tgg gag	gat agg ttt caa cag gag ccc aaa caa ccg
arg arg glu leu lys thr pro trp trp glu	asp arg phe gln gln glu pro lys gln pro
2161/721	2191/731
ccc gca cca gaa ccg cga ccg gcg ccg ccg	ccc gcc aaa ccc gcg cca ccg gcg ggc ccg
pro ala pro glu pro arg pro ala pro pro	pro ala lys pro ala pro pro ala gly pro
2221/741	2251/751
gtt gat gac gac gtc atc tac ccg ccg atg	ctc tcc gag atg gtg ggt gac ccg cac gag
val asp asp asp val ile tyr arg arg met	leu ser glu met val gly asp pro his glu
2281/761	2311/771
ctg gcc cac agc ccc gat ctg gac tgg aag	tcg gtg tgg gac cac ggc tgg tcg gcg gcc
leu ala his ser pro asp leu asp trp lys	ser val trp asp his gly trp ser ala ala
2341/781	2371/791
gcc gag gcc gcg gac aag ccc gtg cag tcc	cgc acg gac tac ggc ctg ccg gtg cgc gaa
ala glu ala ala asp lys pro val gln ser	arg thr asp tyr gly leu pro val arg glu
2401/801	2431/811
ccc ggg gcc ccg tta gtg ccg ggg gcg gcg	gtg cct gag gga ccc gat ccg gag cat ccg
pro gly ala arg leu val pro gly ala ala	val pro glu gly pro asp arg glu his pro
2461/821	2491/831
ggt gca gcg cta gca tcc aac ggc gga ctt	cat ccc ggc cga gcg ccg ccg cac gcg gct
gly ala ala leu ala ser asn gly gly leu	his pro gly arg ala pro arg his ala ala
2521/841	2551/851
gcg gta cgc gac ccc gac gcg gtt cgt gcc	tcc atc agc agc cat ttc ggc ggc gtg cgc
ala val arg asp pro asp ala val arg ala	ser ile ser ser his phe gly gly val arg
2581/861	2611/871
acc ggg ccg tcg cat gcc cgc gag agc agt	cag gga ccc aat cag caa tga
thr gly arg ser his ala arg glu ser ser	gln gly pro asn gln gln OPA

SEQ ID N° 21D (suite)

FIGURE 21D (suite)

84/185

ORF d'après par Cole et al. (Nature 393:537-544) et contenant Rv3365c

```

1/1                                31/11
taa ggg tgc ggc cgg tgg cac ggc cgc ggc cac gtg acc atg ttc gcc cgc ccg acc atc
OCH gly cys gly arg trp his gly arg gly his val thr met phe ala arg pro thr ile
61/21                                91/31
ccg gtc gcg gcg gcc gct tct gat att tcc gcc ccg gct caa ccg gcc cgc ggc aaa cct
pro val ala ala ala ala ser asp ile ser ala pro ala gln pro ala arg gly lys pro
121/41                                151/51
cag caa cgc ccg ccg tcc tgg tcg ccg cgc aac tgg ccg gtc cga tgg aaa gtg ttc acg
gln gln arg pro pro ser trp ser pro arg asn trp pro val arg trp lys val phe thr
181/61                                211/71
atc gcg ctt ctg ccg ctg gta gtg gcg atg gtg tta gca gga ttg ccg gtc gag gct gcg
ile ala leu leu pro leu val val ala met val leu ala gly leu arg val glu ala ala
241/81                                271/91
atg gcc agc acc agc ggc ctg ccg ctg gtc gcc gcg cgc gcc gaa atg ata ccc gcg atc
met ala ser thr ser gly leu arg leu val ala ala arg ala glu met ile pro ala ile
301/101                                331/111
acg aaa tac atg tcg gcg ctg gac gtc gcc gtg ctg gcc agc tcg acc gga cac gat gtg
thr lys tyr met ser ala leu asp val ala val leu ala ser ser thr gly his asp val
361/121                                391/131
gag ggg gcg cag aaa aac ttc acc gcc cgc aag tac gag ctg cag acg cga ctg gcc gac
glu gly ala gln lys asn phe thr ala arg lys tyr glu leu gln thr arg leu ala asp
421/141                                451/151
acc gac gtc atc gca gac gtg ccg tcg gga gtg aac acg ctg ctc aac ggc ggt cag gcg
thr asp val ile ala asp val arg ser gly val asn thr leu leu asn gly gly gln ala
481/161                                511/171
ctg ctg gat aag gtg ctg gcc gac agc atc ggc ttg ccg gat ccg gtc acc gcc tac gcg
leu leu asp lys val leu ala asp ser ile gly leu arg asp arg val thr ala tyr ala
541/181                                571/191
ccg ctg ctg ttg acg gcc cag aac gtg att gac gcg tcg gtg ccg gtt gac agc gag caa
pro leu leu leu thr ala gln asn val ile asp ala ser val arg val asp ser glu gln
601/201                                631/211
atc cga acc cag gtg cag ggt ttg agc cga gcc gtt ggc gcc cgc ggg cag atg acg atg
ile arg thr gln val gln gly leu ser arg ala val gly ala arg gly gln met thr met
661/221                                691/231
cag gag atc ctg gtg act cgc ggc gcc gac ctt gcc gag ccg caa ctg cgc agc gcg atg
gln glu ile leu val thr arg gly ala asp leu ala glu pro gln leu arg ser ala met
721/241                                751/251
gtt acc ctg gcc ggc acc gaa ccc tcg acg ctg ttc ggg atg agc gcg gcg ctc ggt gca
val thr leu ala gly thr glu pro ser thr leu phe gly met ser ala ala leu gly ala
781/261                                811/271
ggc tcg ccg gac acc aag aac ctg cag cag caa atg gtg acc agg atg gcg atc atg tcc
gly ser pro asp thr lys asn leu gln gln gln met val thr arg met ala ile met ser
841/281                                871/291
gat ccg gcc gtt gca ctg gtc aac aac cca gag ctg ctg cac tcg ata cag atc acc cgc
asp pro ala val ala leu val asn asn pro glu leu leu his ser ile gln ile thr arg

```

SEQ ID N° 21F

FIGURE 21F

FEUILLE DE REMPLACEMENT (REGLE 26)

85/185

901/301	931/311
gac att gcc gag cag gtg atc acc gac acc	acc gag gcg gtg acg aag tcg gtg caa agc
asp ile ala glu gln val ile thr asp thr	thr glu ala val thr lys ser val gln ser
961/321	991/331
cag gcc acc gac cgg cgg gat gcc gcg att	cgc gac gcc gtg ctg gtg ttg gcc gcc atc
gln ala thr asp arg arg asp ala ala ile	arg asp ala val leu val leu ala ala ile
1021/341	1051/351
gcg acc gcg atc gtc gtc gtg ttg gtg gtg	gcg cgc acg ctg gtc ggg ccg atg cgg gta
ala thr ala ile val val val leu val val	ala arg thr leu val gly pro met arg val
1081/361	1111/371
ctg cgt gat ggg gcg ctc aag gtt gct cat	acc gat ctc gac ggc gag atc gcg gcg gtc
leu arg asp gly ala leu lys val ala his	thr asp leu asp gly glu ile ala ala val
1141/381	1171/391
cgc gcc ggc gac gag ccg atc ccc gag cca	ctg gcg gtg tac acc acc gag gaa atc ggt
arg ala gly asp glu pro ile pro glu pro	leu ala val tyr thr thr glu glu ile gly
1201/401	1231/411
cag gtc gcg cat gcg gtc gac gag ctg cac	acc cgg gcc ctg ttg ctg gcc ggc gag gaa
gln val ala his ala val asp glu leu his	thr arg ala leu leu leu ala gly glu glu
1261/421	1291/431
acg cgg ttg cga ctg ctg gtc aac gag atg	ttt gag acc atg tcg cgg cgt agc cgt tcc
thr arg leu arg leu leu val asn glu met	phe glu thr met ser arg arg ser arg ser
1321/441	1351/451
ctg gtc gac cag cag ctg tcg gtc atc gac	caa ctg gag cgc aac gag gag gat ccc gcc
leu val asp gln gln leu ser val ile asp	gln leu glu arg asn glu glu asp pro ala
1381/461	1411/471
cga ctc gac agc ctt ttc cgg ctc gat cac	ctg gcc gcc cgg ctg cgc cgc aac agc gcc
arg leu asp ser leu phe arg leu asp his	leu ala ala arg leu arg arg asn ser ala
1441/481	1471/491
aac ctg ctg gtg ctg gcc ggt gcg cag att	acc cgt gac cac cgc gag ccg gtg ccg ctg
asn leu leu val leu ala gly ala gln ile	thr arg asp his arg glu pro val pro leu
1501/501	1531/511
tca acc gtg atc agc gcc gcc gtg tca gag	gtc gag gac tat cgc cgc gtc gac atc gcg
ser thr val ile ser ala ala val ser glu	val glu asp tyr arg arg val asp ile ala
1561/521	1591/531
agg gta ccc gac tgt gcg gta gtc ggc gca	gcg gct ggt ggc gtc att cat ctg ctt gcc
arg val pro asp cys ala val val gly ala	ala ala gly gly val ile his leu leu ala
1621/541	1651/551
gag ctg atc gac aac gcg ttg cgc tac tcg	tca ccg acc aca ccc gtt cgg gtt gcc gcc
glu leu ile asp asn ala leu arg tyr ser	ser pro thr thr pro val arg val ala ala
1681/561	1711/571
gca atc ggc agc gaa ggc agt gtt ctg ctg	cga atc tcg gat tcc ggc ctg ggc atg acc
ala ile gly ser glu gly ser val leu leu	arg ile ser asp ser gly leu gly met thr
1741/581	1771/591
gat gcc gat cgg cgg atg gcc aat atg cgg	ctg cgg gcc ggc ggt gag gtc acc ccg gat
asp ala asp arg arg met ala asn met arg	leu arg ala gly gly glu val thr pro asp
1801/601	1831/611
agt gcc cgg cac atg ggt ctg ttc gta gtc	ggc cgg ctg gcc ggt cgg cac ggc atc cga
ser ala arg his met gly leu phe val val	gly arg leu ala gly arg his gly ile arg

SEQ ID N° 21F (suite 1)

FIGURE 21F (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

86/185

1861/621
 gtc ggg ctg cgc ggt ccg gtg acc ggt gaa
 val gly leu arg gly pro val thr gly glu
 1921/641
 ctg ccg cta gcc gtg ctc gag ggg acg gcc
 leu pro leu ala val leu glu gly thr ala
 1981/661
 gcg atc aag ccg ccg tgt cct gaa ccc gcg
 ala ile lys pro pro cys pro glu pro ala
 2041/681
 atc ggg ccg cta cca ccg gtc acg ttg ctc
 ile gly pro leu pro pro val thr leu leu
 2101/701
 gcc gac gtc ccg gcc cag ccg atg cag cag
 ala asp val pro ala gln pro met gln gln
 2161/721
 gag gat agg ttt caa cag gag ccc aaa caa
 glu asp arg phe gln gln glu pro lys gln
 2221/741
 ccg ccc gcc aaa ccc gcg cca ccg gcg gcc
 pro pro ala lys pro ala pro pro ala gly
 2281/761
 atg ctc tcc gag atg gtg ggt gac ccg cac
 met leu ser glu met val gly asp pro his
 2341/781
 aag tcg gtg tgg gac cac ggc tgg tcg gcg
 lys ser val trp asp his gly trp ser ala
 2401/801
 tcc cgc acg gac tac ggc ctg ccg gtg cgc
 ser arg thr asp tyr gly leu pro val arg
 2461/821
 gcg gtg cct gag gga ccc gat ccg gag cat
 ala val pro glu gly pro asp arg glu his
 2521/841
 ctt cat ccc ggc cga gcg ccg ccg cac gcg
 leu his pro gly arg ala pro arg his ala
 2581/861
 gcc tcc atc agc agc cat ttc ggc ggc gtg
 ala ser ile ser ser his phe gly gly val
 2641/881
 agt cag gga ccc aat cag caa tga
 ser gln gly pro asn gln gln OPA

1891/631
 cag ggc acc ggc acc acc gcc gag gtc tac
 gln gly thr gly thr thr ala glu val tyr
 1951/651
 cca gcg cag ccg cca aag ccg ccg gta ttt
 pro ala gln pro pro lys pro arg val phe
 2011/671
 gcg gcc gat ccg acg gac gtt ccc gcc gcc
 ala ala asp pro thr asp val pro ala ala
 2071/691
 ccg cgc cgt acc ccg ggg tcc agt ggc atc
 pro arg arg thr pro gly ser ser gly ile
 2131/711
 ccg ccg cgc gag ctg aaa aca ccc tgg tgg
 arg arg arg glu leu lys thr pro trp trp
 2191/731
 ccg ccc gca cca gaa ccg cga ccg gcg ccg
 pro pro ala pro glu pro arg pro ala pro
 2251/751
 ccg gtt gat gac gac gtc atc tac ccg ccg
 pro val asp asp asp val ile tyr arg arg
 2311/771
 gag ctg gcc cac agc ccc gat ctg gac tgg
 glu leu ala his ser pro asp leu asp trp
 2371/791
 gcc gcc gag gcc gcg gac aag ccc gtg cag
 ala ala glu ala ala asp lys pro val gln
 2431/811
 gaa ccc ggg gcc ccg tta gtg ccg ggg gcg
 glu pro gly ala arg leu val pro gly ala
 2491/831
 ccg ggt gca gcg cta gca tcc aac ggc gga
 pro gly ala ala leu ala ser asn gly gly
 2551/851
 gct gcg gta cgc gac ccc gac gcg gtt cgt
 ala ala val arg asp pro asp ala val arg
 2611/871
 cgc acc ggg ccg tcg cat gcc cgc gag agc
 arg thr gly arg ser his ala arg glu ser

SEQ ID N° 21F (suite 2)

FIGURE 21F (suite 2)

87/185

31/11
 CTA CGA CAA GGC AAA GGA GCA CAG GGT GAA GCG TGG ACT GAC GGT CGC GGT AGC CGG AGC
 leu arg gln gly lys gly ala gln gly glu ala trp thr asp gly arg gly ser arg ser
 61/21
 CGC CAT TCT GGT CGC AGG TCT TTC CGG ATG TTC AAG CAA CAA GTC GAC TAC AGG AAG CGG
 arg his ser gly arg arg ser phe arg met phe lys gln gln val asp tyr arg lys arg
 121/41
 TGA GAC CAC GAC CGC GGC AGG CAC GAC GGC AAG CCC CGG CGC CGC ATC CGG GCC GAA GGT
 OPA asp his asp arg gly arg his asp gly lys pro arg arg arg ile arg ala glu gly
 181/61
 CGT CAT CGA CGG TAA GGA CCA GAA CGT CAC CGG GTC TGT GGT GTG CAC AAC CGC GGC CGG
 arg his arg arg OCH gly pro glu arg his arg val cys gly val his asn arg gly arg
 241/81
 CAA TGT CAA CAT CGC GAT CGG CGG GGC GGC GAC CGG CAT TGC CGC CGT GCT CAC CGA CGG
 gln cys gln his arg asp arg arg gly gly asp arg his cys arg arg ala his arg arg
 301/101
 CAA CCC TCC GGA GGT GAA GTC CGT TGG GCT CGG TAA CGT CAA CGG CGT CAC GCT GGG ATA
 gln pro ser gly gly glu val arg trp ala arg OCH arg gln arg arg his ala gly ile
 361/121
 CAC GTC GGG CAC CGG ACA GGG TAA CGC TCG GCA ACC AAG GAC GGC AGC CAC TAC AAG ATC
 his val gly his arg thr gly OCH arg ser ala thr lys asp gly ser his tyr lys ile
 391/131

SEQ ID N° 22A

FIGURE 22A

32/11
 TAC GAC AAG GCA AAG GAG CAC AGG GTG AAG CGT GGA CTG ACG GTC GCG GTA GCC GGA GCC
 tyr asp lys ala lys glu his arg val lys arg gly leu thr val ala val ala gly ala
 62/21
 GCC ATT CTG GTC GCA GGT CTT TCC GGA TGT TCA AGC AAC AAG TCG ACT ACA GGA AGC GGT
 ala ile leu val ala gly leu ser gly cys ser ser asn lys ser thr thr gly ser gly
 122/41
 GAG ACC ACG ACC GCG GCA GGC ACG ACG GCA AGC CCC GGC GCC GCA TCC GGG CCG AAG GTC
 glu thr thr thr ala ala gly thr thr ala ser pro gly ala ala ser gly pro lys val
 182/61
 GTC ATC GAC GGT AAG GAC CAG AAC GTC ACC GGG TCT GTG GTG TGC ACA ACC GCG GCC GGC
 val ile asp gly lys asp gln asn val thr gly ser val val cys thr thr ala ala gly
 242/81
 AAT GTC AAC ATC GCG ATC GGC GGG GCG GCG ACC GGC ATT GCC GCC GTG CTC ACC GAC GGC
 asn val asn ile ala ile gly gly ala ala thr gly ile ala ala val leu thr asp gly
 302/101
 AAC CCT CCG GAG GTG AAG TCC GTT GGG CTC GGT AAC GTC AAC GGC GTC ACG CTG GGA TAC
 asn pro pro glu val lys ser val gly leu gly asn val asn gly val thr leu gly tyr
 362/121
 ACG TCG GGC ACC GGA CAG GGT AAC GCT CGG CAA CCA AGG ACG GCA GCC ACT ACA AGA TC
 thr ser gly thr gly gln gly asn ala arg gln pro arg thr ala ala thr thr arg
 392/131

SEQ ID N° 22B

FIGURE 22B

88/185

33/11
 ACG ACA AGG CAA AGG AGC ACA GGG TGA AGC GTG GAC TGA CGG TCG CGG TAG CCG GAG CCG
 thr thr arg gln arg ser thr gly OPA ser val asp OPA arg ser arg AMB pro glu pro
 63/21
 CCA TTC TGG TCG CAG GTC TTT CCG GAT GTT CAA GCA ACA AGT CGA CTA CAG GAA GCG GTG
 pro phe trp ser gln val phe pro asp val gln ala thr ser arg leu gln glu ala val
 123/41
 AGA CCA CGA CCG CGG CAG GCA CGA CGG CAA GCC CCG GCG CCG CAT CCG GGC CGA AGG TCG
 arg pro arg pro arg gln ala arg arg gln ala pro ala pro his pro gly arg arg ser
 183/61
 TCA TCG ACG GTA AGG ACC AGA ACG TCA CCG GGT CTG TGG TGT GCA CAA CCG CGG CCG GCA
 ser ser thr val arg thr arg thr ser pro gly leu trp cys ala gln pro arg pro ala
 243/81
 ATG TCA ACA TCG CGA TCG GCG GGG CGG CGA CCG GCA TTG CCG CCG TGC TCA CCG ACG GCA
 met ser thr ser arg ser ala gly arg arg pro ala leu pro pro cys ser pro thr ala
 303/101
 ACC CTC CGG AGG TGA AGT CCG TTG GGC TCG GTA ACG TCA ACG GCG TCA CGC TGG GAT ACA
 thr leu arg arg OPA ser pro leu gly ser val thr ser thr ala ser arg trp asp thr
 363/121
 CGT CGG GCA CCG GAC AGG GTA ACG CTC GGC AAC CAA GGA CGG CAG CCA CTA CAA GAT C
 arg arg ala pro asp arg val thr leu gly asn gln gly arg gln pro leu gln asp
 393/131

SEQ ID N° 22C

FIGURE 22C

31/11
 GCA CAA CCG CGG CCG GCA ATG TCA ACA TCG CGA TCG GCG GGG CGG CGA CCG GCA TTG CCG
 ala gln pro arg pro ala met ser thr ser arg ser ala gly arg arg pro ala leu pro
 61/21
 CCG TGC TCA CCG ACG GCA ACC CTC CGG AGG TGA AGT CCG TTG GGC TCG GTA ACG TCA ACG
 pro cys ser pro thr ala thr leu arg arg OPA ser pro leu gly ser val thr ser thr
 121/41
 GCG TCA CGC TGG GAT ACA CGT CGG GCA CCG GAC AGG GTA ACG CCT CGG CAA CCA AGG ACG
 ala ser arg trp asp thr arg arg ala pro asp arg val thr pro arg gln pro arg thr
 181/61
 GCA GCC ACT ACA AGA TCA CAG GGT GAA GCG TGG ACT GAC GGT CGC GGT AGC CGG AGC CGC
 ala ala thr thr arg ser gln gly glu ala trp thr asp gly arg gly ser arg ser arg
 241/81
 CAT TCT GGT CGC AGG TCT TTC CGG ATG TTC AAG CAA CAA GTC GAC TAC AGG AAG CGG TGA
 his ser gly arg arg ser phe arg met phe lys gln gln val asp tyr arg lys arg OPA
 301/101
 GAC CAC GAC CGC GGC AGG CAC GAC GGC AAG CCC CGG CGC CGC TCC GGG CCG AAG GTC GTC
 asp his asp arg gly arg his asp gly lys pro arg arg arg ser gly pro lys val val
 361/121
 ATC GAC GGT AAG GAC CAG AAC GTC ACC GGC TCC GTG GTG TGC ACA ACC GCG GCC GGC AAT
 ile asp gly lys asp gln asn val thr gly ser val val cys thr thr ala ala gly asn
 421/141
 GTC AAC ATC GCG ATC GGC GGG GCG GCG ACC GGC ATT GCC GCC GTG CTC ACC GAC GGC AAC
 val asn ile ala ile gly gly ala ala thr gly ile ala ala val leu thr asp gly asn
 481/161
 CCT CCG GAG GTG AAG TCC GTT GGG CTC GGT AAC GTC AAC GGC GTC ACG CTG GGA TAC ACG
 pro pro glu val lys ser val gly leu gly asn val asn gly val thr leu gly tyr thr
 541/181
 TCG GGC ACC GGA CAG GGT AAC GCC TCG GCA ACC AAG GAC GGC AGC CAC TAC AAG ATC
 ser gly thr gly gln gly asn ala ser ala thr lys asp gly ser his tyr lys ile

SEQ ID N° 23A

89/185

32/11
CAC AAC CGC GGC CGG CAA TGT CAA CAT CGC GAT CGG CGG GGC GGC GAC CGG CAT TGC CGC
his asn arg gly arg gln cys gln his arg asp arg arg gly gly asp arg his cys arg
62/21
CGT GCT CAC CGA CGG CAA CCC TCC GGA GGT GAA GTC CGT TGG GCT CGG TAA CGT CAA CGG
arg ala his arg arg gln pro ser gly gly glu val arg trp ala arg OCH arg gln arg
122/41
CGT CAC GCT GGG ATA CAC GTC GGG CAC CGG ACA GGG TAA CGC CTC GGC AAC CAA GGA CGG
arg his ala gly ile his val gly his arg thr gly OCH arg leu gly asn gln gly arg
182/61
CAG CCA CTA CAA GAT CAC AGG GTG AAG CGT GGA CTG ACG GTC GCG GTA GCC GGA GCC GCC
gln pro leu gln asp his arg val lys arg gly leu thr val ala val ala gly ala ala
242/81
ATT CTG GTC GCA GGT CTT TCC GGA TGT TCA AGC AAC AAG TCG ACT ACA GGA AGC GGT GAG
ile leu val ala gly leu ser gly cys ser ser asn lys ser thr thr gly ser gly glu
302/101
ACC ACG ACC GCG GCA GGC ACG ACG GCA AGC CCC GGC GCC GCT CCG GGC CGA AGG TCG TCA
thr thr thr ala ala gly thr thr ala ser pro gly ala ala pro gly arg arg ser ser
362/121
TCG ACG GTA AGG ACC AGA ACG TCA CCG GCT CCG TGG TGT GCA CAA CCG CGG CCG GCA ATG
ser thr val arg thr arg thr ser pro ala pro trp cys ala gln pro arg pro ala met
422/141
TCA ACA TCG CGA TCG GCG GGG CGG CGA CCG GCA TTG CCG CCG TGC TCA CCG ACG GCA ACC
ser thr ser arg ser ala gly arg arg pro ala leu pro pro cys ser pro thr ala thr
482/161
CTC CGG AGG TGA AGT CCG TTG GGC TCG GTA ACG TCA ACG GCG TCA CGC TGG GAT ACA CGT
leu arg arg OPA ser pro leu gly ser val thr ser thr ala ser arg trp asp thr arg
542/181
CGG GCA CCG GAC AGG GTA ACG CCT CGG CAA CCA AGG ACG GCA GCC ACT ACA AGA TC
arg ala pro asp arg val thr pro arg gln pro arg thr ala ala thr thr arg

SEQ ID N° 23B

FIGURE 23B

90/185

33/11
 ACA ACC GCG GCC GGC AAT GTC AAC ATC GCG ATC GGC GGG GCG GCG ACC GGC ATT GCC GCC
 thr thr ala ala gly asn val asn ile ala ile gly gly ala ala thr gly ile ala ala
 63/21
 GTG CTC ACC GAC GGC AAC CCT CCG GAG GTG AAG TCC GTT GGG CTC GGT AAC GTC AAC GGC
 val leu thr asp gly asn pro pro glu val lys ser val gly leu gly asn val asn gly
 123/41
 GTC ACG CTG GGA TAC ACG TCG GGC ACC GGA CAG GGT AAC GCC TCG GCA ACC AAG GAC GGC
 val thr leu gly tyr thr ser gly thr gly gln gly asn ala ser ala thr lys asp gly
 183/61
 AGC CAC TAC AAG ATC ACA GGG TGA AGC GTG GAC TGA CCG TCG CGG TAG CCG CAG CCG CCA
 ser his tyr lys ile thr gly OPA ser val asp OPA arg ser arg AMB pro glu pro pro
 243/81
 TTC TGG TCG CAG GTC TTT CCG GAT GTT CAA GCA ACA AGT CGA CTA CAG GAA GCG GTG AGA
 phe trp ser gln val phe pro asp val gln ala thr ser arg leu gln glu ala val arg
 303/101
 CCA CGA CCG CGG CAG GCA CGA CGG CAA GCC CCG GCG CCG CTC CGG GCC GAA GGT CGT CAT
 pro arg pro arg gln ala arg arg gln ala pro ala pro leu arg ala glu gly arg his
 363/121
 CGA CGG TAA GGA CCA GAA CGT CAC CGG CTC CGT GGT GTG CAC AAC CGC GGC CGG CAA TGT
 arg arg OCH gly pro glu arg his arg leu arg gly val his asn arg gly arg gln cys
 423/141
 CAA CAT CGC GAT CGG CGG GGC GGC GAC CGG CAT TGC CGC CGT GCT CAC CGA CGG CAA CCC
 gln his arg asp arg arg gly gly asp arg his cys arg arg ala his arg arg gln pro
 483/161
 TCC GGA GGT GAA GTC CGT TGG GCT CGG TAA CGT CAA CGG CGT CAC GCT GGG ATA CAC GTC
 ser gly gly glu val arg trp ala arg OCH arg gln arg arg his ala gly ile his val
 543/181
 GGG CAC CGG ACA GGG TAA CGC CTC GGC AAC CAA GGA CGG CAG CCA CTA CAA GAT C
 gly his arg thr gly OCH arg leu gly asn gln gly arg gln pro leu gln asp

SEQ ID N° 23C

FIGURE 23C

91/185

31/11
 CTA ACG ACA GGC AAA GGA GCA CAG GGT GAA GCG TGG ACT GAC GGT CGC GGT AGC CGG AGC
 leu thr thr gly lys gly ala gln gly glu ala trp thr asp gly arg gly ser arg ser
 61/21
 CGC CAT TCT GGT CGC AGG TCT TTC CGG ATG TTC AAG CAA CAA GTC GAC TAC AGG AAG CGG
 arg his ser gly arg arg ser phe arg met phe lys gln gln val asp tyr arg lys arg
 121/41
 TGA GAC CAC GAC CGC GGC AGG CAC GAC GGC AAG CCC CGG CGC CGC TCC GGG CCG AAG GTC
 OPA asp his asp arg gly arg his asp gly lys pro arg arg arg ser gly pro lys val
 181/61
 GTC ATC GAC GGT AAG GAC CAG AAC GTC ACC GGC TCC GTG GTG TGC ACA ACC GCG GCC GGC
 val ile asp gly lys asp gln asn val thr gly ser val val cys thr thr ala ala gly
 241/81
 AAT GTC AAC ATC GCG ATC GGC GGG GCG GCG ACC GGC ATT GCC GCC GTG CTC ACC GAC GGC
 asn val asn ile ala ile gly gly ala ala thr gly ile ala ala val leu thr asp gly
 301/101
 AAC CCT CCG GAG GTG AAG TCC GTT GGG CTC GGT AAC GTC AAC GGC GTC ACG CTG GGA TAC
 asn pro pro glu val lys ser val gly leu gly asn val asn gly val thr leu gly tyr
 361/121
 ACG TCG GGC ACC GGA CAG GGT AAC GCC TCG GCA ACC AAG GAC GGC AGC CAC TAC AAG ATC
 thr ser gly thr gly gln gly asn ala ser ala thr lys asp gly ser his tyr lys ile
 391/131

SEQ ID N° 24A

FIGURE 24A

32/11
 TAA CGA CAG GCA AAG GAG CAC AGG GTG AAG CGT GGA CTG ACG GTC GCG GTA GCC GGA GCC
 OCH arg gln ala lys glu his arg val lys arg gly leu thr val ala val ala gly ala
 62/21
 GCC ATT CTG GTC GCA GGT CTT TCC GGA TGT TCA AGC AAC AAG TCG ACT ACA GGA AGC GGT
 ala ile leu val ala gly leu ser gly cys ser ser asn lys ser thr thr qly ser gly
 122/41
 GAG ACC ACG ACC GCG GCA GGC ACG ACG GCA AGC CCC GGC GCC GCT CCG GGC CGA AGG TCG
 glu thr thr thr ala ala gly thr thr ala ser pro gly ala ala pro gly arg arg ser
 182/61
 TCA TCG ACG GTA AGG ACC AGA ACG TCA CCG GCT CCG TGG TGT GCA CAA CCG CGG CCG GCA
 ser ser thr val arg thr arg thr ser pro ala pro trp cys ala gln pro arg pro ala
 242/81
 ATG TCA ACA TCG CGA TCG GCG GGG CGG CGA CCG GCA TTG CCG CCG TGC TCA CCG ACG GCA
 met ser thr ser arg ser ala gly arg arg pro ala leu pro pro cys ser pro thr ala
 302/101
 ACC CTC CCG AGG TGA AGT CCG TTG GGC TCG GTA ACG TCA ACG GCG TCA CGC TGG GAT ACA
 thr leu arg arg OPA ser pro leu gly ser val thr ser thr ala ser arg trp asp thr
 362/121
 CGT CGG GCA CCG GAC AGG GTA ACG CCT CGG CAA CCA AGG ACG GCA GCC ACT ACA AGA TC
 arg arg ala pro asp arg val thr pro arg gln pro arg thr ala ala thr thr arg
 392/131

SEQ ID N° 24B

FIGURE 24B

FEUILLE DE REMPLACEMENT (REGLE 26)

92/185

33/11
 AAC GAC AGG CAA AGG AGC ACA GGG TGA AGC GTG GAC TGA CGG TCG CGG TAG CCG GAG CCG
 asn asp arg gln arg ser thr gly OPA ser val asp OPA arg ser arg AMB pro glu pro
 63/21
 93/31
 CCA TTC TGG TCG CAG GTC TTT CCG GAT GTT CAA GCA ACA AGT CGA CTA CAG GAA GCG GTG
 pro phe trp ser gln val phe pro asp val gln ala thr ser arg leu gln glu ala val
 123/41
 153/51
 AGA CCA CGA CCG CGG CAG GCA CGA CGG CAA GCC CCG GCG CCG CTC CGG GCC GAA GGT CGT
 arg pro arg pro arg gln ala arg arg gln ala pro ala pro leu arg ala glu gly arg
 183/61
 213/71
 CAT CGA CGG TAA GGA CCA GAA CGT CAC CGG CTC CGT GGT GTG CAC AAC CGC GGC CGG CAA
 his arg arg OCH gly pro glu arg his arg leu arg gly val his asn arg gly arg gln
 243/81
 273/91
 TGT CAA CAT CGC GAT CGG CGG GGC GGC GAC CGG CAT TGC CGC CGT GCT CAC CGA CGG CAA
 cys gln his arg asp arg arg gly qly asp arg his cys arg arg ala his arg arg gln
 303/101
 333/111
 CCC TCC GGA GGT GAA GTC CGT TGG GCT CGG TAA CGT CAA CGG CGT CAC GCT GGG ATA CAC
 pro ser gly gly glu val arg trp ala arg OCH arg gln arg arg his ala gly ile his
 363/121
 393/131
 GTC GGG CAC CGG ACA GGG TAA CGC CTC GGC AAC CAA GGA CGG CAG CCA CTA CAA GAT C
 val gly his arg thr gly OCH arg leu gly asn gln gly arg gln pro leu gln asp

SEQ ID N° 24C

FIGURE 24C

Amorce directe

5' ACG CGG CGC AGC CTG TTG 3'

SEQ ID N° 25

FIGURE 25

Amorce inverse

5' CGA CCT TGG GAT TCG CCT 3'

SEQ ID N° 26

FIGURE 26

93/185

31/11
 CCT ACC AGC AAG AGC CCA GGG CTT CAC AGG ACC TAA AAG GAG TAG CGC CCA TGG GCT TGA
 pro thr ser lys ser pro gly leu his arg thr OCH lys glu AMB arg pro trp ala OPA
 61/21
 TCC AAT TTT CCT TCC GCC CCG TGC AAT ACC ATC TGC AAG ACC AGC GAC GGC CCG TGG TTG
 ser asn phe pro ser ala pro cys asn thr ile cys lys thr ser asp gly pro trp leu
 121/41
 CGG TCG CGC AGC TTG CGG AAA CGG GGT ATG GAC CCT GCC GTA CCG TTG TTG CCA CTT GAT
 arg ser arg ser leu arg lys arg gly met asp pro ala val pro leu leu pro leu asp
 181/61
 GTC GTC GCT CTC CAC CCG TCG GGG GGC GAA AGC CAT TCC GAC ACT GGG ATC CTC AAA ACG
 val val ala leu his pro ser gly gly glu ser his ser asp thr gly ile leu lys thr
 241/81
 TCG GCT GAG TGT CTG CAG GGC TCC GGG GAG CAG CCG ATC ATC ACC ATG TAC GAA CTG AAT
 ser ala glu cys leu gln gly ser gly glu gln pro ile ile thr met tyr glu leu asn
 301/101
 AAG TCC CCC CCG CGC GAC TTC CAG ACA TTT GTT GTG GTT TCG GTT GAG GCC GAG GCG AGG
 lys ser pro pro arg asp phe gln thr phe val val val ser val glu ala glu ala arg
 361/121
 CTC ATT TCG CAG CAA GCG GTC TCC GGG TCG CAG CAT CGT TGC GGC GAT CGC GGC GCA GTC
 leu ile ser gln gln ala val ser gly ser gln his arg cys gly asp arg gly ala val
 421/141
 GTC GGA CGA GTC GTC GTC AAC GAC CAC GAT C
 val gly arg val val val asn asp his asp

SEQ ID N° 27A

FIGURE 27A

31/11
 CTA CCA GCA AGA GCC CAG GGC TTC ACA GGA CCT AAA AGG AGT AGC GCC CAT GGG CTT GAT
 leu pro ala arg ala gln gly phe thr gly pro lys arg ser ser ala his gly leu asp
 61/21
 CCA ATT TTC CTT CCG CCC CGT GCA ATA CCA TCT GCA AGA CCA GCG ACG GCC CGT GGT TGC
 pro ile phe leu pro pro arg ala ile pro ser ala arg pro ala thr ala arg gly cys
 121/41
 GGT CGC GCA GCT TGC GGA AAC GGG GTA TGG ACC CTG CCG TAC CGT TGT TGC CAC TTG ATG
 gly arg ala ala cys gly asn gly val trp thr leu pro tyr arg cys cys his leu met
 181/61
 TCG TCG CTC TCC ACC CGT CGG GGG GCG AAA GCC ATT CCG ACA CTG GGA TCC TCA AAA CGT
 ser ser leu ser thr arg arg gly ala lys ala ile pro thr leu gly ser ser lys arg
 241/81
 CGG CTG AGT GTC TGC AGG GCT CCG GGG AGC AGC CGA TCA TCA CCA TGT ACG AAC TGA ATA
 arg leu ser val cys arg ala pro gly ser ser arg ser ser pro cys thr asn OPA ile
 301/101
 AGT CCC CCC CGC GCG ACT TCC AGA CAT TTG TTG TGG TTT CGG TTG AGG CCG AGG CGA GGC
 ser pro pro arg ala thr ser arg his leu leu trp phe arg leu arg pro arg arg gly
 361/121
 TCA TTT CGC AGC AAG CGG TCT CCG GGT CGC AGC ATC GTT GCG GCG ATC GCG GCG CAG TCG
 ser phe arg ser lys arg ser pro gly arg ser ile val ala ala ile ala ala gln ser
 421/141
 TCG GAC GAG TCG TCG TCA ACG ACC ACG ATC
 ser asp glu ser ser ser thr thr thr ile

SEQ ID N° 27B

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 27B

94/185

33/11
TAC CAG CAA GAG CCC AGG GCT TCA CAG GAC CTA AAA GGA GTA GCG CCC ATG GGC TTG ATC
tyr gln gln glu pro arg ala ser gln asp leu lys gly val ala pro met gly leu ile
63/21
CAA TTT TCC TTC CGC CCC GTG CAA TAC CAT CTG CAA GAC CAG CGA CGG CCC GTG GTT GCG
gln phe ser phe arg pro val gln tyr his leu gln asp gln arg arg pro val val ala
123/41
GTC GCG CAG CTT GCG GAA ACG GGG TAT GGA CCC TGC CGT ACC GTT GTT GCC ACT TGA TGT
val ala gln leu ala glu thr gly tyr gly pro cys arg thr val val ala thr OPA cys
183/61
CGT CGC TCT CCA CCC GTC GGG GGG CGA AAG CCA TTC CGA CAC TGG GAT CCT CAA AAC GTC
arg arg ser pro pro val gly gly arg lys pro phe arg his trp asp pro gln asn val
243/81
GGC TGA GTG TCT GCA GGG CTC CGG GGA GCA GCC GAT CAT CAC CAT GTA CGA ACT GAA TAA
gly OPA val ser ala gly leu arg gly ala ala asp his his his val arg thr glu OCH
303/101
GTC CCC CCC GCG CGA CTT CCA GAC ATT TGT TGT GGT TTC GGT TGA GGC CGA GGC GAG GCT
val pro pro ala arg leu pro asp ile cys cys gly phe gly OPA gly arg gly glu ala
363/121
CAT TTC GCA GCA AGC GGT CTC CGG GTC GCA GCA TCG TTG CGG CGA TCG CGG CGC AGT CGT
his phe ala ala ser gly leu arg val ala ala ser leu arg arg ser arg arg ser arg
423/141
CGG ACG AGT CGT CGT CAA CGA CCA CGA TC
arg thr ser arg arg gln arg pro arg

SEQ ID N° 27C

FIGURE 27C

MKTGTATRRRLAVLIALALPGAVALLAEPSATGASDPCAASEVAR
TVGSVAKSMGDYLDSPETNQVMTAVLQQQVGPVSVASLKAHFEANPK
VASDLHALSQPLTDLSTRCSLPISGLQAIGLMQAVQGARR

SEQ ID N° 28

FIGURE 28

GTGGGCAAGC	AGCTAGCCGC	GCTCGCCGCG	CTGGTCGGTG	CGTGCATGCT	CGCAGCCGGA	60
TGCACCAACG	TGGTCGACGG	GACCGCCGTG	GCTGCCGACA	AATCCGGACC	ACTGCATCAG	120
GATCCGATAC	CGGTTTCAGC	GCTTGAAGGG	CTGCTTCTCG	ACTTGAGCCA	GATCAATGCC	180
GCGCTGGGTG	CGACATCGAT	GAAGGTGTGG	TTCAACGCCA	AGGCAATGTG	GGACTGGAGC	240
AAGAGCGTGG	CCGACAAGAA	TTGCCTGGCT	ATCGACGGTC	CAGCACAGGA	AAAGGTCTAT	300
GCCGGCACCG	GGTGGACCGC	TATGCGCGGC	CAACGGCTGG	ATGACAGCAT	CGATGACTCC	360
AAGAAACGCG	ACCACTACGC	CATTCAAGCG	GTCGTCGGCT	TCCCGACCGC	ACATGATGCC	420
GAGGAGTTCT	ACAGCTCCTC	GGTGCAAAGC	TGGAGCAGCT	GCTCGAACCG	CCGGTTTGTC	480
GAAGTCACCC	CCGGACAGGA	CGACGCCGCC	TGGACTGTGG	CTGACGTTGT	CAACGACAAC	540
GGCATGCTCA	GTAGCTCGCA	GGTTCAGGAA	GCGGCGGACG	GATGGACCTG	CCAGCGTGCC	600
CTGACTGCGC	GCAACAACGT	CACATCGAC	ATTGTACAGT	GCGCCTATAG	CCAACCGGAT	660
TTGGTGGCGA	TTGGCATCGC	TAACCAAATC	GCGGCCAAGG	TTGCTAAGCA	GTAG	714

SEQ ID N° 29

FEUILLE DE REMPLACEMENT (REGLE 26)

95/185

MGKQLAALAALVGACMLAAGCTNVVDGTAVAADKSGPLHQDPIPVFTSALEGLLLDLSQINAALGATS
 MKVWFNAKAMWDWSKSVADKNCLAI DGPAQEKVYAGTGFTWTAMRGQRLDDSIDDSKKRDHYAIQAVV
 GFPTAHDAEEFYSSSVQSWSSCSNRRFEVETFTPGQDDAAWTVADVVDNGMLSSSQVQEGGDGWTCC
 RALTARNNVITIDIVTCAYSQPDLVFTAIGIANQIAAKVAKQ

SEQ ID N° 30

FIGURE 30

1/1	31/11
AGG CGA ATA CCC GCG AGG GCA GCG CGA CGG	CGG CCC TGC CGG CGC CGT GGC TGC TGA ACA
arg arg ile pro ala arg ala ala arg arg	arg pro cys arg arg arg gly cys OPA thr
61/21	91/31
ACA CAT CCC AGC CGC GCA CGC TTC CGG TAT	GCG GCA GGA TAA ACG ACC CCA ACA GCA CGA
thr his pro ser arg ala arg phe arg tyr	ala ala gly OCH thr thr pro thr ala arg
121/41	151/51
ACA CCA GGA TTG CGA CAA CCA AAG CCC TCG	CGC CTG GCT CGA TTT CGC GCG CAA CGC GGC
thr pro gly leu arg gln pro lys pro ser	arg leu ala arg phe arg ala gln arg gly
181/61	211/71
GTT CTG CCG CCT CGA TCT CAG CGC GGA GGG	CGT CGA GAT C
val leu pro pro arg ser gln arg gly gly	arg arg asp

SEQ ID N° 31A

FIGURE 31A

1/1	31/11
GGC GAA TAC CCG CGA GGG CAG CGC GAC GGC	GGC CCT GCC GGC GCC GTG GCT GCT GAA CAA
gly glu tyr pro arg gly gln arg asp gly	gly pro ala gly ala val ala ala glu gln
61/21	91/31
CAC ATC CCA GCC GCG CAC GCT TCC GGT ATG	CGG CAG GAT AAA CGA CCC CAA CAG CAC GAA
his ile pro ala ala his ala ser gly met	arg gln asp lys arg pro gln gln his glu
121/41	151/51
CAC CAG GAT TGC GAC AAC CAA AGC CCT CGC	GCC TGG CTC GAT TTC GCG CGC AAC GCG GCG
his gln asp cys asp asn gln ser pro arg	ala trp leu asp phe ala arg asn ala ala
181/61	211/71
TTC TGC CGC CTC GAT CTC AGC GCG GAG GGC	GTC GAG ATC
phe cys arg leu asp leu ser ala glu gly	val glu ile

SEQ ID N° 31B

FIGURE 31B

96/185

1/1 31/11
 GCG AAT ACC CGC GAG GGC AGC GCG ACG GCG GCC CTG CCG GCG CCG TGG CTG CTG AAC AAC
 ala asn thr arg glu gly ser ala thr ala ala leu pro ala pro trp leu leu asn asn
 61/21 91/31
 ACA TCC CAG CCG CGC ACG CTT CCG GTA TGC GGC AGG ATA AAC GAC CCC AAC AGC ACG AAC
 thr ser gln pro arg thr leu pro val cys gly arg ile asn asp pro asn ser thr asn
 121/41 151/51
 ACC AGG ATT GCG ACA ACC AAA GCC CTC GCG CCT GGC TCG ATT TCG CGC GCA ACG CGG CGT
 thr arg ile ala thr thr lys ala leu ala pro gly ser ile ser arg ala thr arg arg
 181/61 211/71
 TCT GCC GCC TCG ATC TCA GCG CGG AGG GCG TCG AGA TC
 ser ala ala ser ile ser ala arg arg ala ser arg

SEQ ID N° 31C

FIGURE 31C

ORF d'après Cole et al. (Nature 393:537-544) et contenant seq31A

1/1 31/11
 taa acg acc cca aca gca cga aca cca gga ttg cga caa cca' aag ccc tcg cgc ctg gct
 OCH thr thr pro thr ala arg thr pro gly leu arg gln pro lys pro ser arg leu ala
 61/21 91/31
 cga ttt cgc gcg caa cgc ggc gtt ctg ccg cct cga tct cag cgc gga ggg cgt cga gat
 arg phe arg ala gln arg gly val leu pro pro arg ser gln arg gly gly arg arg asp
 121/41 151/51
 ccc cgg cgt cgt gtt cgt ggc tca tca tct gca tcc tcc ggg ctt ggc cgc gct gac cgg
 pro arg arg arg val arg gly ser ser ser ala ser ser gly leu gly arg ala asp arg
 181/61 211/71
 cag ccc gac ccc agg cat gcc cag gcc gac ggc gcg ccc cgg ctg ccc ggc ggt gtg cgc
 gln pro asp pro arg his ala gln ala asp gly ala pro arg leu pro gly gly val arg
 241/81 271/91
 gtc gcc ggc gcg ggt gcg gcg gtg ggt cag gac gcc ggc gtc ggc gat gag gtg gtg cgg
 val ala gly ala gly ala ala val gly gln asp ala gly val gly asp glu val val arg
 301/101 331/111
 cgc cgc ttc ggt gac ctt cgt ggt gat gac gtc gcc ggg acg cac gcg cgg ctg gcc ggc
 arg arg phe gly asp leu arg gly asp asp val ala gly thr his ala arg leu ala gly
 361/121 391/131
 ggt gaa gtg cac cag gcg ccc gtc gcg cgc ccg ccc gct cat gcg cgc cgt gac ggt gtc
 gly glu val his gln ala pro val ala arg pro pro ala his ala arg arg asp gly val
 421/141 451/151
 ctt gcg ccc ttc ccc ggt ggc cac cag cac ctc gac ggc ctg ccc gac cag ggc gcg gtt
 leu ala pro phe pro gly gly his gln his leu asp gly leu pro asp gln gly ala val
 481/161 511/171
 ggc ttc cag cga gat ttg ctc ctg cag cgc gat cag gcg ttc ata gcg ttc ctg cac aac
 gly phe gln arg asp leu leu leu gln arg asp gln ala phe ile ala phe leu his asn
 541/181 571/191
 ggc ttt cgg cag ctg tcc gtc gag ttg cgc ggc cgg tgt ccc ggg ccg ctt gga gta ttg
 gly phe arg gln leu ser val glu leu arg gly arg cys pro gly pro leu gly val leu
 601/201 631/211
 gaa ggt aaa tgc ggc cgc gaa gcg ggc ccg gcg cac cac gtc gag cgt ggc cgc gaa gtc
 glu gly lys cys gly arg glu ala gly pro ala his his val glu arg gly arg glu val
 661/221 691/231
 ctc ttc ggt ctc ccc ggg gaa acc gac gat cag atc ggt ggt aat cgc ggc atg cgg gat
 leu phe gly leu pro gly glu thr asp asp gln ile gly gly asn arg gly met arg asp
 721/241 751/251
 ggc cgc ccg cac gcg ctc gat gat gcc gag gta gcg ctc ggc acg ata gga ccg ccg cat
 gly arg pro his ala leu asp asp ala glu val ala leu gly thr ile gly pro pro his
 781/261 811/271
 cgc gcg cag gat ccg gtc gga tcc gga ctg tag
 arg ala gln asp pro val gly ser gly leu AMB

SEQ ID N° 31F

FEUILLE DE REMPLACEMENT (REGLE 26)

97/185

```

1/1                                31/11
aga ctg gtg tac acg gag acc aag ctg aac tcg gca ttc tcc ttc ggc ggg cct aag tgt
arg leu val tyr thr glu thr lys leu asn ser ala phe ser phe gly gly pro lys cys
61/21                                91/31
cta gtg aag gtc att cag aaa ctg tcg ggc ttg agc atc aac cgg ttc atc gcg att gac
leu val lys val ile gln lys leu ser gly leu ser ile asn arg phe ile ala ile asp
121/41                                151/51
ttc gtc ggt ttc gcg cgg atg gtc gag gcc ctc ggc ggc gtc gag gta tgc agc acc acc
phe val gly phe ala arg met val glu ala leu gly gly val glu val cys ser thr thr
181/61                                211/71
ccg ttg cgg gac tac gaa ctg ggc acg gtg ctg gag cac gcc gga cgc cag gtc att gac
pro leu arg asp tyr glu leu gly thr val leu glu his ala gly arg gln val ile asp
241/81                                271/91
ggg ccg acc gcg ctg aac tat gtg cgc gct cgc cag gtc acc acc gag agc aat ggc gac
gly pro thr ala leu asn tyr val arg ala arg gln val thr thr glu ser asn gly asp
301/101                                331/111
tac ggg cgc atc aaa cgc cag cag ttg ttt ttg tcg tcg ctg ctg cgt tcg atg atc
tyr gly arg ile lys arg gln gln leu phe leu ser ser leu leu arg ser met ile

```

SEQ ID N° 32A

FIGURE 32A

```

1/1                                31/11
gac tgg tgt aca cgg aga cca agc tga act cgg cat tct cct tcg gcg ggc cta agt gtc
asp trp cys thr arg arg pro ser OPA thr arg his ser pro ser ala gly leu ser val
61/21                                91/31
tag tga agg tca ttc aga aac tgt cgg gct tga gca tca acc ggt tca tcg cga ttg act
AMB OPA arg ser phe arg asn cys arg ala OPA ala ser thr gly ser ser arg leu thr
121/41                                151/51
tcg tcg gtt tcg cgc gga tgg tcg agg ccc tcg gcg gcg tcg agg tat gca gca cca ccc
ser ser val ser arg gly trp ser arg pro ser ala ala ser arg tyr ala ala pro pro
181/61                                211/71
cgt tgc ggg act acg aac tgg gca cgg tgc tgg agc acg ccg gac gcc agg tca ttg acg
arg cys gly thr thr asn trp ala arg cys trp ser thr pro asp ala arg ser leu thr
241/81                                271/91
ggc cga ccg cgc tga act atg tgc gcg ctc gcc agg tca cca ccg aga gca atg gcg act
gly arg pro arg OPA thr met cys ala leu ala arg ser pro pro arg ala met ala thr
301/101                                331/111
acg ggc gca tca aac gcc agc agt tgt ttt tgt cgt cgc tgc tgc gtt cga tga tc
thr gly ala ser asn ala ser ser cys phe cys arg arg cys cys val arg OPA

```

SEQ ID N° 32B

FIGURE 32B

98/185

```

1/1                               31/11
act ggt gta cac gga gac caa gct gaa ctc ggc att ctc ctt cgg cgg gcc taa gtg tct
thr gly val his gly asp gln ala glu leu gly ile leu leu arg arg ala OCH val ser
61/21                               91/31
agt gaa ggt cat tca gaa act gtc ggg ctt gag cat caa ccg gtt cat cgc gat tga ctt
ser glu gly his ser glu thr val gly leu glu his gln pro val his arg asp OPA leu
121/41                               151/51
cgt cgg ttt cgc gcg gat ggt cga ggc cct cgg cgg cgt cga ggt atg cag cac cac ccc
arg arg phe arg ala asp gly arg gly pro arg arg arg arg gly met gln his his pro
181/61                               211/71
gtt gcg gga cta cga act ggg cac ggt gct gga gca cgc cgg acg cca ggt cat tga cgg
val ala gly leu arg thr gly his gly ala gly ala arg arg thr pro gly his OPA arg
241/81                               271/91
gcc gac cgc gct gaa cta tgt gcg cgc tcg cca ggt cac cac cga gag caa tgg cga cta
ala asp arg ala glu leu cys ala arg ser pro gly his his arg glu gln trp arg leu
301/101                               331/111
cgg gcg cat caa acg cca gca gtt gtt ttt gtc gtc gct gct gcg ttc gat gat c
arg ala his gln thr pro ala val val phe val val ala ala ala phe asp asp

```

SEQ ID N° 32C

FIGURE 32C

séquence Rv0822c prédite par Cole et al. (Nature 393:537-544) et contenant seq 32A

```

1/1                               31/11
atg agt gac ggc gag agc gcc gcg ccg tgg gca cgg ctc tcc gag tca gca ttc ccc gat
Met ser asp gly glu ser ala ala pro trp ala arg leu ser glu ser ala phe pro asp
61/21                               91/31
ggt gtt gac cga tgg atc acg gta ccg ccc gcc aca tgg gtg gca gcc cag ggt ccg cgg
gly val asp arg trp ile thr val pro pro ala thr trp val ala ala gln gly pro arg
121/41                               151/51
gac acc cag aat gtc ggc tgt cat gcc acc ggc gcc gtt agt gtg gcc gat ctg atc gcc
asp thr gln asn val gly cys his ala thr gly ala val ser val ala asp leu ile ala
181/61                               211/71
agg ctc ggc ccc gct ttt cct gac ctc ccc acg cac cgc cat gtc gcc ccc gaa ccc gag
arg leu gly pro ala phe pro asp leu pro thr his arg his val ala pro glu pro glu
241/81                               271/91
cca tcc ggc cgc ggc ccg aag gtc cac gac gac gcc gac gac cag cag gac acc gag gct
pro ser gly arg gly pro lys val his asp asp ala asp asp gln gln asp thr glu ala
301/101                               331/111
atc gcc atc ccg gcc cac tcg ctc gag ttc ctc tcg gag ctt ccc gac ctc cgg gca gcc
ile ala ile pro ala his ser leu glu phe leu ser glu leu pro asp leu arg ala ala
361/121                               391/131
aac tat ccg cgc gcc gac cac gcc cgc cgt gaa ccc gag cta ccc ggc aag cag cta acc
asn tyr pro arg ala asp his ala arg arg glu pro glu leu pro gly lys gln leu thr
421/141                               451/151
gga tcg gct cga gtg cgg cca ttg cgg atc cgc cga acg tcg ccc gcg ccc gcc aag cca
gly ser ala arg val arg pro leu arg ile arg arg thr ser pro ala pro ala lys pro
481/161                               511/171
gcg ccg aac tcc ggc cgg cgc ccg atg gtg ctc gcc gcg cgc tcg ctg gcg gct ctg ttt
ala pro asn ser gly arg arg pro met val leu ala ala arg ser leu ala ala leu phe
541/181                               571/191
gcc gct ctg gcg ttg gcg ctg acc ggc ggg gca tgg cag tgg agc gcg tcg aag aac agc
ala ala leu ala leu ala leu thr gly gly ala trp gln trp ser ala ser lys asn ser
601/201                               631/211
cgg ctg aac atg gta agc gcg ctc gac ccg cat tcg ggc gac atc gtc aac ccc agc ggg
arg leu asn met val ser ala leu asp pro his ser gly asp ile val asn pro ser gly

```

SEQ ID N° 32D

FEUILLE DE REMPLACEMENT (REGLE 26)

99/185

661/221
 cag cat ggc gac gag aac ttc ttg ctc gtc ggt atg gac tct cgt gcc ggg gcg aac gcc
 gln his gly asp glu asn phe leu leu val gly met asp ser arg ala gly ala asn ala
 721/241
 aat atc ggc gcc ggc gac gcc gag gac gcc ggc ggc gca cgt tcg gac acc gtc atg ctg
 asn ile gly ala gly asp ala glu asp ala gly gly ala arg ser asp thr val met leu
 781/261
 gtc aac att ccg gcc agc cgc gag cgg gtc gtc gcg gtg tcg ttc ccc cgc gac ctg gcg
 val asn ile pro ala ser arg glu arg val val ala val ser phe pro arg asp leu ala
 841/281
 atc act cca atc caa tgc gag gcg tgg aac ccc gag acc ggt aag tac gga ccc atc tac
 ile thr pro ile gln cys glu ala trp asn pro glu thr gly lys tyr gly pro ile tyr
 901/301
 gag gag aag acg gga acg atg ggt ccc aga ctg gtg tac acg gag acc aag ctg aac tcg
 asp glu lys thr gly thr met gly pro arg leu val tyr thr glu thr lys leu asn ser
 961/321
 gca ttc tcc ttc ggc ggg cct aag tgt cta gtc aag gtc att cag aaa ctg tcg ggc ttg
 ala phe ser phe gly gly pro lys cys leu val lys val ile gln lys leu ser gly leu
 1021/341
 agc atc aac cgg ttc atc gcg att gac ttc gtc ggt ttc gcg cgg atg gtc gag gcc ctc
 ser ile asn arg phe ile ala ile asp phe val gly phe ala arg met val glu ala leu
 1081/361
 ggc ggc gtc gag gta tgc agc acc acc ccg ttg cgg gac tac gaa ctg ggc acg gtg ctg
 gly gly val glu val cys ser thr thr pro leu arg asp tyr glu leu gly thr val leu
 1141/381
 gag cac gcc gga cgc cag gtc att gac ggg ccg acc gcg ctg aac tat gtg cgc gct cgc
 glu his ala gly arg gln val ile asp gly pro thr ala leu asn tyr val arg ala arg
 1201/401
 cag gtc acc acc gag agc aat ggc gac tac ggg cgc atc aaa cgc cag cag ttg ttt ttg
 gln val thr thr glu ser asn gly asp tyr gly arg ile lys arg gln gln leu phe leu
 1261/421
 tcg tcg ctg ctg cgt tcg atg atc tcg acg gac acc ttg ttc aac ctc agc agg ctc aac
 ser ser leu leu arg ser met ile ser thr asp thr leu phe asn leu ser arg leu asn
 1321/441
 aac gtc gtc aac atg ttc atc ggt aac agc tac gtg gac aac gtc aag acc aaa gac ctg
 asn val val asn met phe ile gly asn ser tyr val asp asn val lys thr lys asp leu
 1381/461
 gtc gaa ctc ggt cga tcg ttg cag cat atg gcg gcc ggg cac gtc acg ttc gtg acc gtt
 val glu leu gly arg ser leu gln his met ala ala gly his val thr phe val thr val
 1441/481
 ccg acc ggt ata acc gac cag aac ggc gac gag ccc ccg cgt acc tcc gac atg aag gcg
 pro thr gly ile thr asp gln asn gly asp glu pro pro arg thr ser asp met lys ala
 1501/501
 ctt ttc acc gcc atc atc gac gac gat ccg ctg ccc ctg gaa aac gat cac aac gcc cag
 leu phe thr ala ile ile asp asp asp pro leu pro leu glu asn asp his asn ala gln
 1561/521
 cgt ctg ggc aac acg ccg tcg acc ccg ccg acc acc acc aag aag gcg ccg cag gcg ggt
 arg leu gly asn thr pro ser thr pro pro thr thr thr lys lys ala pro gln ala gly
 1621/541
 ctg acc aac gag att cag cac cag cag gtt acg acg acc tcg cca aaa gag gtc aca gtg
 leu thr asn glu ile gln his gln gln val thr thr thr ser pro lys glu val thr val
 1681/561
 cag gtc tct aac tcg acc ggc cag gcc ggt ttg gcc acc acc gcc acc gat cag ctc aag
 gln val ser asn ser thr gly gln ala gly leu ala thr thr ala thr asp gln leu lys
 1741/581
 cgg aac ggc ttc aac gtg atg gct ccg gac gac tac ccg agt tcg ctg ctg gcc acc aca
 arg asn gly phe asn val met ala pro asp asp tyr pro ser ser leu leu ala thr thr
 1801/601
 gtg ttt ttt tcg ccc ggc aac gaa cag gct gcc gcc acc gtg gcc gcc gtg ttc ggc cag
 val phe phe ser pro gly asn glu gln ala ala ala thr val ala ala val phe gly gln
 1861/621
 tca aag atc gag cgg gtg acc ggg atc ggc caa ctg gtc cag gtg gtg ctg ggc caa gac
 ser lys ile glu arg val thr gly ile gly gln leu val gln val val leu gly gln asp
 691/231
 751/251
 811/271
 871/291
 931/311
 991/331
 1051/351
 1111/371
 1171/391
 1231/411
 1291/431
 1351/451
 1411/471
 1471/491
 1531/511
 1591/531
 1651/551
 1711/571
 1771/591
 1831/611
 1891/631

SEQ ID N° 32D (suite 1)

 FIGURE 32D (suite 1)
 FEUILLE DE REMPLACEMENT (REGLE 26)

100/185

1921/641
 ttc agc gcg gtg cgc gct ccc ctg ccg agt ggc tcc acc gtc agc gtg cag ata agc cgc
 phe ser ala val arg ala pro leu pro ser gly ser thr val ser val gln ile ser arg
 1981/661
 aac tcc tcc agc cca ccg acc aag ctg ccc gag gac ctg acg gtc acc aac gcc gcc gac
 asn ser ser ser pro pro thr lys leu pro glu asp leu thr val thr asn ala ala asp
 2041/681
 acc acc tgc gag tag
 thr thr cys glu AMB

1951/651

2011/671

SEQ ID N° 32D (suite 2)
 FIGURE 32D (suite 2)

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv0822c

1/1
 tag gac atg agt gac ggc gag agc gcc gcg ccc tgg gca cgg ctc tcc gag tca gca ttc
 AMB asp met ser asp gly glu ser ala ala pro trp ala arg leu ser glu ser ala phe
 61/21
 ccc gat ggt gtt gac cga tgg atc acg gta ccg ccc gcc aca tgg gtg gca gcc cag ggt
 pro asp gly val asp arg trp ile thr val pro pro ala thr trp val ala ala gln gly
 121/41
 ccg cgg gac acc cag aat gtc ggc tgt cat gcc acc ggc gcc gtt agt gtg gcc gat ctg
 pro arg asp thr gln asn val gly cys his ala thr gly ala val ser val ala asp leu
 181/61
 atc gcc agg ctc ggc ccc gct ttt cct gac ctc ccc acg cac cgc cat gtc gcc ccc gaa
 ile ala arg leu gly pro ala phe pro asp leu pro thr his arg his val ala pro glu
 241/81
 ccc gag cca tcc ggc cgc ggc ccg aag gtc cac gac gac gcc gac gac cag cag gac acc
 pro glu pro ser gly arg gly pro lys val his asp asp ala asp asp gln gln asp thr
 301/101
 gag gct atc gcc atc ccg gcc cac tgc ctc gag ttc ctc tgc gag ctt ccc gac ctc cgg
 glu ala ile ala ile pro ala his ser leu glu phe leu ser glu leu pro asp leu arg
 361/121
 gca gcc aac tat ccg cgc gcc gac cac gcc cgc cgt gaa ccc gag cta ccc ggc aag cag
 ala ala asn tyr pro arg ala asp his ala arg arg glu pro glu leu pro gly lys gln
 421/141
 cta acc gga tgc gct cga gtg cgg cca ttg cgg atc cgc cga acg tgc ccc gcg ccc gcc
 leu thr gly ser ala arg val arg pro leu arg ile arg arg thr ser pro ala pro ala
 481/161
 aag cca gcg ccg aac tcc ggc cgg cgc ccg atg gtg ctg gcc gcg cgc tgc ctg gcg gct
 lys pro ala pro asn ser gly arg arg pro met val leu ala ala arg ser leu ala ala
 541/181
 ctg ttt gcc gct ctg gcg ttg gcg ctg acc ggc ggg gca tgg cag tgg agc gcg tgc aag
 leu phe ala ala leu ala leu ala leu thr gly gly ala trp gln trp ser ala ser lys
 601/201
 aac agc cgg ctg aac atg gta agc gcg ctc gac ccg cat tgc ggc gac atc gtc aac ccc
 asn ser arg leu asn met val ser ala leu asp pro his ser gly asp ile val asn pro
 661/221
 agc ggg cag cat ggc gac gag aac ttc ttg ctc gtc ggt atg gac tct cgt gcc ggg gcg
 ser gly gln his gly asp glu asn phe leu leu val gly met asp ser arg ala gly ala
 721/241
 aac gcc aat atc ggc gcc ggc gac gcc gag gac gcc ggc ggc gca cgt tgc gac acc gtc
 asn ala asn ile gly ala gly asp ala glu asp ala gly gly ala arg ser asp thr val
 781/261
 atg ctg gtc aac att ccg gcc agc cgc gag ccg gtc gtc gcg gtg tgc ttc ccc cgc gac
 met leu val asn ile pro ala ser arg glu arg val val ala val ser phe pro arg asp

SEQ ID N° 32F

FEUILLE DE REMPLACEMENT (REGLE 26)
 FIGURE 32F

101/185

841/281
 ctg gcg atc act cca atc caa tgc gag gcg tgg aac ccc gag acc ggt aag tac gga ccc
 leu ala ile thr pro ile gln cys glu ala trp asn pro glu thr gly lys tyr gly pro
 901/301
 atc tac gac gag aag acg gga acg atg ggt ccc aga ctg gtg tac acg gag acc aag ctg
 ile tyr asp glu lys thr gly thr met gly pro arg leu val tyr thr glu thr lys leu
 961/321
 aac tcg gca ttc tcc ttc ggc ggg cct aag tgt cta gtg aag gtc att cag aaa ctg tcg
 asn ser ala phe ser phe gly gly pro lys cys leu val lys val ile gln lys leu ser
 1021/341
 ggc ttg agc atc aac cgg ttc atc gcg att gac ttc gtc ggt ttc gcg cgg atg gtc gag
 gly leu ser ile asn arg phe ile ala ile asp phe val gly phe ala arg met val glu
 1081/361
 gcc ctg ggc ggc gtc gag gta tgc agc acc acc ccg ttg cgg gac tac gaa ctg ggc acg
 ala leu gly gly val glu val cys ser thr thr pro leu arg asp tyr glu leu gly thr
 1141/381
 gtg ctg gag cac gcc gga cgc cag gtc att gac ggg ccg acc gcg ctg aac tat gtg cgc
 val leu glu his ala gly arg gln val ile asp gly pro thr ala leu asn tyr val arg
 1201/401
 gct cgc cag gtc acc acc gag agc aat ggc gac tac ggg cgc atc aaa cgc cag cag ttg
 ala arg gln val thr thr glu ser asn gly asp tyr gly arg ile lys arg gln gln leu
 1261/421
 ttt ttg tcg tcg ctg ctg cgt tcg atg atc tcg acg gac acc ttg ttc aac ctg agc agg
 phe leu ser ser leu leu arg ser met ile ser thr asp thr leu phe asn leu ser arg
 1321/441
 ctg aac aac gtc gtc aac atg ttc atc ggt aac agc tac gtg gac aac gtc aag acc aaa
 leu asn asn val val asn met phe ile gly asn ser tyr val asp asn val lys thr lys
 1381/461
 gac ctg gtc gaa ctg ggt cga tcg ttg cag cat atg gcg gcc ggg cac gtc acg ttc gtg
 asp leu val glu leu gly arg ser leu gln his met ala ala gly his val thr phe val
 1441/481
 acc gtt ccg acc ggt ata acc gac cag aac ggc gac gag ccc ccg cgt acc tcc gac atg
 thr val pro thr gly ile thr asp gln asn gly asp glu pro pro arg thr ser asp met
 1501/501
 aag gcg ctt ttc acc gcc atc atc gac gac gat ccg ctg ccc ctg gaa aac gat cac aac
 lys ala leu phe thr ala ile ile asp asp asp pro leu pro leu glu asn asp his asn
 1561/521
 gcc cag cgt ctg ggc aac acg ccg tcg acc ccg ccg acc acc acc aag aag gcg ccg cag
 ala gln arg leu gly asn thr pro ser thr pro pro thr thr thr lys lys ala pro gln
 1621/541
 gcg ggt ctg acc aac gag att cag cac cag cag gtt acg acg acc tcg cca aaa gag gtc
 ala gly leu thr asn glu ile gln his gln gln val thr thr thr ser pro lys glu val
 1681/561
 aca gtg cag gtc tct aac tcg acc ggc cag gcc ggt ttg gcc acc acc gcc acc gat cag
 thr val gln val ser asn ser thr gly gln ala gly leu ala thr thr ala thr asp gln
 1741/581
 ctg aag cgg aac ggc ttc aac gtg atg gct ccg gac gac tac ccg agt tcg ctg ctg gcc
 leu lys arg asn gly phe asn val met ala pro asp asp tyr pro ser ser leu leu ala
 1801/601
 acc aca gtg ttt ttt tcg ccc ggc aac gaa cag gct gcc gcc acc gtg gcc gcc gtg ttc
 thr thr val phe phe ser pro gly asn glu gln ala ala ala thr val ala ala val phe
 1861/621
 ggc cag tca aag atc gag cgg gtg acc ggg atc ggc caa ctg gtc cag gtg gtg ctg ggc
 gly gln ser lys ile glu arg val thr gly ile gly gln leu val gln val val leu gly
 1921/641
 caa gac ttc agc gcg gtg cgc gct ccc ctg ccg agt ggc tcc acc gtc agc gtg cag ata
 gln asp phe ser ala val arg ala pro leu pro ser gly ser thr val ser val gln ile
 1981/661
 agc cgc aac tcc tcc agc cca ccg acc aag ctg ccc gag gac ctg acg gtc acc aac gcc
 ser arg asn ser ser ser pro pro thr lys leu pro glu asp leu thr val thr asn ala
 2041/681
 gcc gac acc acc tgc gag tag
 ala asp thr thr cys glu AMB

SEQ ID 32F (suite 1)

 FIGURE 32F (suite 1)
 FEUILLE DE REMPLACEMENT (REGLE 26)

102/185

1/1 31/11
 CGT CAC CTC TGC CAT GGT CCA TCT ACG GTA TCT GCG ACA AGG GCA GCG TCG ATC CCT CGA
 arg his leu cys his gly pro ser thr val ser ala thr arg ala ala ser ile pro arg
 61/21 91/31
 CAT GCA GAG TCG GTG TTC GCT TCA CGC GAA CTA GGC GCG CCT AGC CTG GAC GAG TCC CCG
 his ala glu ser val phe ala ser arg glu leu gly ala pro ser leu asp glu ser pro
 121/41 151/51
 GGC CGA CAT TCG CCC GAG GCC TTG GCC TCC ATC ACC TAA TTG TGT GCA AAA CCG TAT CTA
 gly arg his ser pro glu ala leu ala ser ile thr OCH leu cys ala lys pro tyr leu
 181/61 211/71
 ATT GAT ACG ATT GCG CAC ATG GCT ATC TGG GAT C
 ile asp thr ile ala his met ala ile trp asp

SEQ ID N° 33A

FIGURE 33A

1/1 31/11
 GTC ACC TCT GCC ATG GTC CAT CTA CGG TAT CTG CGA CAA GGG CAG CGT CGA TCC CTC GAC
 val thr ser ala met val his leu arg tyr leu arg gln gly gln arg arg ser leu asp
 61/21 91/31
 ATG CAG AGT CGG TGT TCG CTT CAC GCG AAC TAG GCG CGC CTA GCC TGG ACG AGT CCC CGG
 met gln ser arg cys ser leu his ala asn AMB ala arg leu ala trp thr ser pro arg
 121/41 151/51
 GCC GAC ATT CGC CCG AGG CCT TGG CCT CCA TCA CCT AAT TGT GTG CAA AAC CGT ATC TAA
 ala asp ile arg pro arg pro trp pro pro ser pro asn cys val gln asn arg ile OCH
 181/61 211/71
 TTG ATA CGA TTG CGC ACA TGG CTA TCT GGG ATC
 leu ile arg leu arg thr trp leu ser gly ile

SEQ ID N° 33B

FIGURE 33B

1/1 31/11
 CCG TCA CCT CTG CCA TGG TCC ATC TAC GGT ATC TGC GAC AAG GGC AGC GTC GAT CCC TCG
 pro ser pro leu pro trp ser ile tyr gly ile cys asp lys gly ser val asp pro ser
 61/21 91/31
 ACA TGC AGA GTC GGT GTT CGC TTC ACG CGA ACT AGG CGC GCC TAG CCT GGA CGA GTC CCC
 thr cys arg val gly val arg phe thr arg thr arg arg ala AMB pro gly arg val pro
 121/41 151/51
 GGG CCG ACA TTC GCC CGA GGC CTT GGC CTC CAT CAC CTA ATT GTG TGC AAA ACC GTA TCT
 gly pro thr phe ala arg gly leu gly leu his his leu ile val cys lys thr val ser
 181/61 211/71
 AAT TGA TAC GAT TGC GCA CAT GGC TAT CTG GGA TC
 asn OPA tyr asp cys ala his gly tyr leu gly

SEQ ID N° 33C

FIGURE 33C
 FEUILLE DE REMPLACEMENT (REGLE 26)

103/185

séquence Rv1044 prédite par Cole et al. (Nature 393:537-544) et contenant seq33A

```

1/1                               31/11
ttg tgt gca aaa ccg tat cta att gat acg att gcg cac atg gct atc tgg gat cgc ctc
leu cys ala lys pro tyr leu ile asp thr ile ala his met ala ile trp asp arg leu
61/21                               91/31
gtc gag gtt gcc gcc gag caa cat ggc tac gtc acg act cgc gat gcg cga gac atc ggc
val glu val ala ala glu gln his gly tyr val thr thr arg asp ala arg asp ile gly
121/41                               151/51
gtc gac cct gtg cag ctc cgc ctc cta gcg ggg cgc gga cgt ctt gag cgt gtc ggc cga
val asp pro val gln leu arg leu leu ala gly arg gly arg leu glu arg val gly arg
181/61                               211/71
ggt gtg tac cgg gtg ccc gtg ctg ccg cgt ggt gag cac gac gat ctc gca gcc gca gtg
gly val tyr arg val pro val leu pro arg gly glu his asp asp leu ala ala ala val
241/81                               271/91
tcg tgg act ttg ggg cgt ggc gtt atc tcg cat gag tcg gcc ttg gcg ctt cat gcc ctc
ser trp thr leu gly arg gly val ile ser his glu ser ala leu ala leu his ala leu
301/101                               331/111
gct gac gtg aac ccg tcg cgc atc cat ctc acc gtc ccg cgc aac aac cat ccg cgt gcg
ala asp val asn pro ser arg ile his leu thr val pro arg asn asn his pro arg ala
361/121                               391/131
gcc ggg ggc gag ctg tac cga gtt cac cgc cgc gac ctc cag gca gcc cac gtc act tcg
ala gly gly glu leu tyr arg val his arg arg asp leu gln ala ala his val thr ser
421/141                               451/151
gtc gac gga ata ccc gtc acg acg gtt gcg cgc acc atc aaa gac tgc gtg aag acg ggc
val asp gly ile pro val thr thr val ala arg thr ile lys asp cys val lys thr gly
481/161                               511/171
acg gat cct tat cag ctt cgg gcc gcg atc gag cga gcc gaa gcc gag ggc acg ctt cgt
thr asp pro tyr gln leu arg ala ala ile glu arg ala glu ala glu gly thr leu arg
541/181                               571/191
cgt ggg tca gca gct gag cta cgc gct gcg ctc gat gag acc act gcc gga tta cgc gct
arg gly ser ala ala glu leu arg ala ala leu asp glu thr thr ala gly leu arg ala
601/201
cgg ccg aag cga gca tcg gcg tga
arg pro lys arg ala ser ala OPA

```

SEQ ID N° 33D

FIGURE 33D

104/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv1044

```

1/1                                31/11
taa ttg tgt gca aaa ccg tat cta att gat acg att gcg cac atg gct atc tgg gat cgc
OCH leu cys ala lys pro tyr leu ile asp thr ile ala his met ala ile trp asp arg
61/21                                91/31
ctc gtc gag gtt gcc gcc gag caa cat ggc tac gtc acg act cgc gat gcg cga gac atc
leu val glu val ala ala glu gln his gly tyr val thr thr arg asp ala arg asp ile
121/41                                151/51
ggc gtc gac ect gtg cag ctc cgc ctc cta gcg ggg cgc gga cgt ctt gag cgt gtc ggc
gly val asp pro val gln leu arg leu leu ala gly arg gly arg leu glu arg val gly
181/61                                211/71
cga ggt gtg tac cgg gtg ccc gtg ctg ccg cgt ggt gag cac gac gat ctc gca gcc gca
arg gly val tyr arg val pro val leu pro arg gly glu his asp asp leu ala ala ala
241/81                                271/91
gtg tcg tgg act ttg ggg cgt ggc gtt atc tcg cat gag tcg gcc ttg gcg ctt cat gcc
val ser trp thr leu gly arg gly val ile ser his glu ser ala leu ala leu his ala
301/101                                331/111
ctc gct gac gtg aac ccg tcg cgc atc cat ctc acc gtc ccg cgc aac aac cat ccg cgt
leu ala asp val asn pro ser arg ile his leu thr val pro arg asn asn his pro arg
361/121                                391/131
gcg gcc ggg ggc gag ctg tac cga gtt cac cgc cgc gac ctc cag gca gcc cac gtc act
ala ala gly gly glu leu tyr arg val his arg arg asp leu gln ala ala his val thr
421/141                                451/151
tcg gtc gac gga ata ccc gtc acg acg gtt gcg cgc acc atc aaa gac tgc gtg aag acg
ser val asp gly ile pro val thr thr val ala arg thr ile lys asp cys val lys thr
481/161                                511/171
ggc acg gat cct tat cag ctt cgg gcc gcg atc gag cga gcc gaa gcc gag ggc acg ctt
gly thr asp pro tyr gln leu arg ala ala ile glu arg ala glu ala glu gly thr leu
541/181                                571/191
cgt cgt ggg tca gca gct gag cta cgc gct gcg ctc gat gag acc act gcc gga tta cgc
arg arg gly ser ala ala glu leu arg ala ala leu asp glu thr thr ala gly leu arg
601/201
gct cgg ccg aag cga gca tcg gcg tga
ala arg pro lys arg ala ser ala OPA

```

SEQ ID N° 33F

FIGURE 33F

```

1/1                                31/11
ATC CAA CCT GCT GGG CCT GCG CCT TCG AAT CGA CGG CCA GGC CAC CGC TCG CTG CCG GCA
ile gln pro ala gly pro ala pro ser asn arg arg pro gly his arg ser leu pro ala
61/21                                91/31
ACA ACA CCT GGA ATG GGG ACC TTT TCG GTG TTG CTG GTA ACC GGG ACA ACC GGC ACC ACG
thr thr pro gly met gly thr phe ser val leu leu val thr gly thr thr gly thr thr
121/41                                151/51
CCT CGG TCG AGA CGT ATC GCG GCA GCG TTG GCC CTG TCG TTG CTG ACA ATT ACC GCT GGC
pro arg ser arg arg ile ala ala ala leu ala leu ser leu leu thr ile thr ala gly
181/61                                211/71
CGC CGC ATA TTT GCC GCG CTG CCG CGG GCC GGA TC
arg arg ile phe ala ala leu pro arg ala gly

```

SEQ ID N° 34A

FIGURE 34A
FEUILLE DE REMPLACEMENT (REGLE 26)

105/185

```

1/1                                31/11
TCC AAC CTG CTG GGC CTG CGC CTT CGA ATC GAC GGC CAG GCC ACC GCT CGC TGC CGG CAA
ser asn leu leu gly leu arg leu arg ile asp gly gln ala thr ala arg cys arg gln
61/21                                91/31
CAA CAC CTG GAA TGG GGA CCT TTT CGG TGT TGC TGG TAA CCG GGA CAA CCG GCA CCA CGC
gln his leu glu trp gly pro phe arg cys cys trp OCH pro gly gln pro ala pro arg
121/41                                151/51
CTC GGT CGA GAC GTA TCG CGG CAG CGT TGG CCC TGT CGT TGC TGA CAA TTA CCG CTG GCC
leu gly arg asp val ser arg gln arg trp pro cys arg cys OPA gln leu pro leu ala
181/61                                211/71
GCC GCA TAT TTG CCG CGC TGC CGC GGG CCG GAT C
ala ala tyr leu pro arg cys arg gly pro asp

```

SEQ ID N° 34B

FIGURE 34B

```

1/1                                31/11
GAT CCA ACC TGC TGG GCC TGC GCC TTC GAA TCG ACG GCC AGG CCA CCG CTC GCT GCC GGC
asp pro thr cys trp ala cys ala phe glu ser thr ala arg pro pro leu ala ala gly
61/21                                91/31
AAC AAC ACC TGG AAT GGG GAC CTT TTC GGT GTT GCT GGT AAC CGG GAC AAC CGG CAC CAC
asn asn thr trp asn gly asp leu phe gly val ala gly asn arg asp asn arg his his
121/41                                151/51
GCC TCG GTC GAG ACG TAT CGC GGC AGC GTT GGC CCT GTC GTT GCT GAC AAT TAC CGC TGG
ala ser val glu thr tyr arg gly ser val gly pro val val ala asp asn tyr arg trp
181/61                                211/71
CCG CCG CAT ATT TGC CGC GCT GCC GCG GGC CGG ATC
pro pro his ile cys arg ala ala ala gly arg ile

```

SEQ ID N° 34C

FIGURE 34C

106/185

ORF d'après Cole et al. (Nature 393:537-544) contenant seq34A

```

1/1                               31/11
tag ccg cag ggc cct gcg gct agg cgc ggc cgg tgc cgt tgg ccg cgg cgg caa tcg atg
AMB pro gln gly pro ala ala arg arg gly arg cys arg trp pro arg arg gln ser met
61/21                               91/31
ttg cag cag tta caa cgc caa atg gag tct gag cgc atc gtc gag ttc gat cag ctc ggc
leu gln gln leu gln arg gln met glu ser glu arg ile val glu phe asp gln leu gly
121/41                               151/51
agg gga gac gtt gcg cag cga cgg atc caa cct gct ggg cct gcg cct tcg aat cga cgg
arg gly asp val ala gln arg arg ile gln pro ala gly pro ala pro ser asn arg arg
181/61                               211/71
cca ggc cac cgc tcg ctg ccg gca aca aca cct gga atg ggg acc ttt tcg gtg ttg ctg
pro gly his arg ser leu pro ala thr thr pro gly met gly thr phe ser val leu leu
241/81                               271/91
gta acc ggg aca acc ggc acc acg cct cgg tcg aga cgt atc gcg gca gcg ttg gcc ctg
val thr gly thr thr gly thr thr pro arg ser arg arg ile ala ala ala leu ala leu
301/101                               331/111
tcg ttg ctg aca att acc gct ggc cgc cgc ata ttt gcc gcg ctg ccg cgg gcc gga tcc
ser leu leu thr ile thr ala gly arg arg ile phe ala ala leu pro arg ala gly ser
361/121                               391/131
agg tcg acc tgc cag atc tca ccg cgc agc atc tac gcc gtt cgc tgc aaa ccg ccg act
arg ser thr cys gln ile ser pro arg ser ile tyr ala val arg cys lys pro pro thr
421/141                               451/151
gcg acg gca ggc cca ctc tct tgg cat gcg tcc aat gct gcg acg tcc tcg gta gac aag
ala thr ala gly pro leu ser trp his ala ser asn ala ala thr ser ser val asp lys
481/161                               511/171
ctc acg ctt ggc ttc atg ccg cag tcc tac cca tgt agt aac aga tag
leu thr leu gly phe met pro gln ser tyr pro cys ser asn arg AMB

```

SEQ ID N° 34F

FIGURE 34F

```

1/1                               31/11
CAG TCT GTC GGC AAG GAG GGA CGC ATG CCA CTC TCC GAT CAT GAG CAG CGG ATG CTT GAC
gln ser val gly lys glu gly arg met pro leu ser asp his glu gln arg met leu asp
61/21                               91/31
CAG ATC GAG AGC GCT CTC TAC GCC GAA GAT CCC AAG TTC GCA TCG AGT GTC CGT GGC GGG
gln ile glu ser ala leu tyr ala glu asp pro lys phe ala ser ser val arg gly gly
121/41                               151/51
GGC TTC CGC GCA CCG ACC GCG CGG CGG CGC CTG CAG GGC GCG GCG TTG TTC ATC ATC GGT
gly phe arg ala pro thr ala arg arg arg leu gln gly ala ala leu phe ile ile gly
181/61                               211/71
CTG GGG ATG TTG GTT TCC GGC GTG GCG TTC AAA GAG ACC ATG ATC GGA AGT TTC CCG ATA
leu gly met leu val ser gly val ala phe lys glu thr met ile gly ser phe pro ile
241/81                               271/91
CTC AGC GTT TTC GGT TTT GTC GTG ATG TTC GGT GGT GTG GTG TAT GCC ATC ACC GGT CCT
leu ser val phe gly phe val val met phe gly gly val val tyr ala ile thr gly pro
301/101                               331/111
CGG TTG TCC GGC AGG ATG GAT CGT GGC GGA TCG GCT GCT GGG GCT TCG CGC CAG CGT CGT
arg leu ser gly arg met asp arg gly gly ser ala ala gly ala ser arg gln arg arg
361/121                               391/131
ACC AAG GGG GCC GGG GGC TCA TTC ACC AGC CGT ATG GAA GAT C
thr lys gly ala gly gly ser phe thr ser arg met glu asp

```

SEQ ID N° 35A

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 35A

107/185

```

1/1                                31/11
GAC AGT CTG TCG GCA AGG AGG GAC GCA TGC CAC TCT CCG ATC ATG AGC AGC GGA TGC TTG
asp ser leu ser ala arg arg asp ala cys his ser pro ile met ser ser gly cys leu
61/21                                91/31
ACC AGA TCG AGA GCG CTC TCT ACG CCG AAG ATC CCA AGT TCG CAT CGA GTG TCC GTG GCG
thr arg ser arg ala leu ser thr pro lys ile pro ser ser his arg val ser val ala
121/41                                151/51
GGG GCT TCC GCG CAC CGA CCG CGC GGC GGC GCC TGC AGG GCG CGG CGT TGT TCA TCA TCG
gly ala ser ala his arg pro arg gly gly ala cys arg ala arg arg cys ser ser ser
181/61                                211/71
GTC TGG GGA TGT TGG TTT CCG GCG TGG CGT TCA AAG AGA CCA TGA TCG GAA GTT TCC CGA
val trp gly cys trp phe pro ala trp arg ser lys arg pro OPA ser glu val ser arg
241/81                                271/91
TAC TCA GCG TTT TCG GTT TTG TCG TGA TGT TCG GTG GTG TGG TGT ATG CCA TCA CCG GTC
tyr ser ala phe ser val leu ser OPA cys ser val val trp cys met pro ser pro val
301/101                                331/111
CTC GGT TGT CCG GCA GGA TGG ATC GTG GCG GAT CCG CTG CTG GGG CTT CGC GCC AGC GTC
leu gly cys pro ala gly trp ile val ala asp arg leu leu gly leu arg ala ser val
361/121                                391/131
GTA CCA AGG GGG CCG GGG GCT CAT TCA CCA GCC GTA TGG AAG ATC
val pro arg gly pro gly ala his ser pro ala val trp lys ile

```

SEQ ID N° 35B

FIGURE 35B

```

1/1                                31/11
ACA GTC TGT CGG CAA GGA GGG ACG CAT GCC ACT CTC CGA TCA TGA GCA GCG GAT GCT TGA
thr val cys arg gln gly gly thr his ala thr leu arg ser OPA ala ala asp ala OPA
61/21                                91/31
CCA GAT CGA GAG CGC TCT CTA CGC CGA AGA TCC CAA GTT CGC ATC GAG TGT CCG TGG CGG
pro asp arg glu arg ser leu arg arg arg ser gln val arg ile glu cys pro trp arg
121/41                                151/51
GGG CTT CCG CGC ACC GAC CGC GCG GCG GCG CCT GCA GGG CGC GGC GTT GTT CAT CAT CGG
gly leu pro arg thr asp arg ala ala ala pro ala gly arg gly val val his his arg
181/61                                211/71
TCT GGG GAT GTT GGT TTC CGG CGT GGC GTT CAA AGA GAC CAT GAT CGG AAG TTT CCC GAT
ser gly asp val gly phe arg arg gly val gln arg asp his asp arg lys phe pro asp
241/81                                271/91
ACT CAG CGT TTT CGG TTT TGT CGT GAT GTT CGG TGG TGT GGT GTA TGC CAT CAC CGG TCC
thr gln arg phe arg phe cys arg asp val arg trp cys gly val cys his his arg ser
301/101                                331/111
TCG GTT GTC CGG CAG GAT GGA TCG TGG CGG ATC GGC TGC TGG GGC TTC GCG CCA GCG TCG
ser val val arg gln asp gly ser trp arg ile gly cys trp gly phe ala pro ala ser
361/121                                391/131
TAC CAA GGG GGC CGG GGG CTC ATT CAC CAG CCG TAT GGA AGA TC
tyr gln gly gly arg gly leu ile his gln pro tyr gly arg

```

SEQ ID N° 35C

FIGURE 35C
FEUILLE DE REMPLACEMENT (REGLE 26)

108/185

séquence Rv2169c prédite par Cole et al. (Nature 393:537-544) et contenant partiellement seq35A

1/1	31/11
atg cca ctc tcc gat cat gag cag cgg atg	ctt gac cag atc gag agc gct ctc tac gcc
Met pro leu ser asp his glu gln arg met	leu asp gln ile glu ser ala leu tyr ala
61/21	91/31
gaa gat ccc aag ttc gca tcg agt gtc cgt	ggc ggg ggc ttc cgc gca ccg acc gcg cgg
glu asp pro lys phe ala ser ser val arg	gly gly gly phe arg ala pro thr ala arg
121/41	151/51
cgg cgc ctg cag ggc gcg gcg ttg ttc atc	atc ggt ctg ggg atg ttg gtt tcc ggc gtg
arg arg leu gln gly ala ala leu phe ile	ile gly leu gly met leu val ser gly val
181/61	211/71
gcg ttc aaa gag acc atg atc gga agt ttc	ccg ata ctc agc gtt ttc ggt ttt gtc gtg
ala phe lys glu thr met ile gly ser phe	pro ile leu ser val phe gly phe val val
241/81	271/91
atg ttc ggt ggt gtg gtg tat gcc atc acc	ggt cct cgg ttg tcc ggc agg atg gat cgt
met phe gly gly val val tyr ala ile thr	gly pro arg leu ser gly arg met asp arg
301/101	331/111
ggc gga tcg gct gct ggg gct tcg cgc cag	cgt cgt acc aag ggg gcc ggg ggc tca ttc
gly gly ser ala ala gly ala ser arg gln	arg arg thr lys gly ala gly gly ser phe
361/121	391/131
acc agc cgt atg gaa gat cgg ttc cgg cgc	cgc ttc gac gag taa
thr ser arg met glu asp arg phe arg arg	arg phe asp glu OCH

SEQ ID N° 35D

FIGURE 35D

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv2169c

1/1	31/11
tga cag tct gtc ggc aag gag gga cgc atg	cca ctc tcc gat cat gag cag cgg atg ctt
OPA gln ser val gly lys glu gly arg met	pro leu ser asp his glu gln arg met leu
61/21	91/31
gac cag atc gag agc gct ctc tac gcc gaa	gat ccc aag ttc gca tcg agt gtc cgt ggc
asp gln ile glu ser ala leu tyr ala glu	asp pro lys phe ala ser ser val arg gly
121/41	151/51
ggg ggc ttc cgc gca ccg acc gcg cgg cgg	cgc ctg cag ggc gcg gcg ttg ttc atc atc
gly gly phe arg ala pro thr ala arg arg	arg leu gln gly ala ala leu phe ile ile
181/61	211/71
ggt ctg ggg atg ttg gtt tcc ggc gtg gcg	ttc aaa gag acc atg atc gga agt ttc ccg
gly leu gly met leu val ser gly val ala	phe lys glu thr met ile gly ser phe pro
241/81	271/91
ata ctc agc gtt ttc ggt ttt gtc gtg atg	ttc ggt ggt gtg gtg tat gcc atc acc ggt
ile leu ser val phe gly phe val val met	phe gly gly val val tyr ala ile thr gly
301/101	331/111
cct cgg ttg tcc ggc agg atg gat cgt ggc	gga tcg gct gct ggg gct tcg cgc cag cgt
pro arg leu ser gly arg met asp arg gly	gly ser ala ala gly ala ser arg gln arg
361/121	391/131
cgt acc aag ggg gcc ggg ggc tca ttc acc	agc cgt atg gaa gat cgg ttc cgg cgc cgc
arg thr lys gly ala gly gly ser phe thr	ser arg met glu asp arg phe arg arg arg
421/141	
ttc gac gag taa	
phe asp glu OCH	

SEQ ID 35F

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 35F

109/185

```

1/1                               31/11
GAC CTG GGA CGA AGA CGA CGG CAG CAG CCG CAA TCA GAT CTA CCC GGT CCT GGT CAA CGT
asp leu gly arg arg arg arg gln gln pro gln ser asp leu pro gly pro gly gln arg
61/21                               91/31
CAA TGG ACA CCC GAC TAC GGT GCG CCT GCG CGG CTC GAC AAT GCG CGG TTC CTG TTG CCC
gln trp thr pro asp tyr gly ala pro ala arg leu asp asn ala arg phe leu leu pro
121/41                               151/51
GTG GTC GGA GTG CCA CCC GAC CAG GCC ACC GAC TTC GGC TCC GCT GTT GCA CCA GAA ACG
val val gly val pro pro asp gln ala thr asp phe gly ser ala val ala pro glu thr
181/61                               211/71
ACG GCG CCG GTC TGG ATC ACC ATG CTG TGG CCG CTG GCC GAC CGG CCC CGG TTG GCC CCC
thr ala pro val trp ile thr met leu trp pro leu ala asp arg pro arg leu ala pro
241/81                               271/91
GGG GCA CCC GGT GGC ACC GTT CCC GTC CGG CTG GTC GAC GAC GAC CTG GCA AAC TCG CTG
gly ala pro gly gly thr val pro val arg leu val asp asp asp leu ala asn ser leu
301/101                               331/111
GCC AAC GGC GGC CGG CTG GAC ATC CTC CTG TCG GCG GCC GAG TTC GCC ACC AAC CGG GAA
ala asn gly gly arg leu asp ile leu leu ser ala ala glu phe ala thr asn arg glu
361/121                               391/131
GTC GAC CCC GAC GGC GCC GTC GGC CGA GCG CTG TGC CTG GCC ATC GAC CCA GAT C
val asp pro asp gly ala val gly arg ala leu cys leu ala ile asp pro asp

```

SEQ ID N° 36A

FIGURE 36A

```

1/1                               31/11
ACC TGG GAC GAA GAC GAC GGC AGC AGC CGC AAT CAG ATC TAC CCG GTC CTG GTC AAC GTC
thr trp asp glu asp asp gly ser ser arg asn gln ile tyr pro val leu val asn val
61/21                               91/31
AAT GGA CAC CCG ACT ACG GTG CGC CTG CGC GGC TCG ACA ATG CGC GGT TCC TGT TGC CCG
asn gly his pro thr thr val arg leu arg gly ser thr met arg gly ser cys cys pro
121/41                               151/51
TGG TCG GAG TGC CAC CCG ACC AGG CCA CCG ACT TCG GCT CCG CTG TTG CAC CAG AAA CGA
trp ser glu cys his pro thr arg pro pro thr ser ala pro leu leu his gln lys arg
181/61                               211/71
CGG CGC CGG TCT GGA TCA CCA TGC TGT GGC CGC TGG CCG ACC GGC CCC GGT TGG CCC CCG
arg arg arg ser gly ser pro cys cys gly arg trp pro thr gly pro gly trp pro pro
241/81                               271/91
GGG CAC CCG GTG GCA CCG TTC CCG TCC GGC TGG TCG ACG ACG ACC TGG CAA ACT CGC TGG
gly his pro val ala pro phe pro ser gly trp ser thr thr thr trp gln thr arg trp
301/101                               331/111
CCA ACG GCG GCC GGC TGG ACA TCC TCC TGT CGG CGG CCG AGT TCG CCA CCA ACC GGG AAG
pro thr ala ala gly trp thr ser ser cys arg arg pro ser ser pro pro thr gly lys
361/121                               391/131
TCG ACC CCG ACG GCG CCG TCG GCC GAG CGC TGT GCC TGG CCA TCG ACC CAG ATC
ser thr pro thr ala pro ser ala glu arg cys ala trp pro ser thr gln ile

```

SEQ ID N° 36B

FIGURE 36B
FEUILLE DE REMPLACEMENT (REGLE 26)

110/185

```

1/1                                31/11
CCT GGG ACG AAG ACG ACG GCA GCA GCC GCA ATC AGA TCT ACC CGG TCC TGG TCA ACG TCA
pro gly thr lys thr thr ala ala ala ala ile arg ser thr arg ser trp ser thr ser
61/21                                91/31
ATG GAC ACC CGA CTA CGG TGC GCC TGC GCG GCT CGA CAA TGC GCG GTT CCT GTT GCC CGT
met asp thr arg leu arg cys ala cys ala ala arg gln cys ala val pro val ala arg
121/41                                151/51
GGT CGG AGT GCC ACC CGA CCA GGC CAC CGA CTT CGG CTC CGC TGT TGC ACC AGA AAC GAC
gly arg ser ala thr arg pro gly his arg leu arg leu arg cys cys thr arg asn asp
181/61                                211/71
GGC GCC GGT CTG GAT CAC CAT GCT GTG GCC GCT GGC CGA CCG GCC CCG GTT GGC CCC CGG
gly ala gly leu asp his his ala val ala ala gly arg pro ala pro val gly pro arg
241/81                                271/91
GGC ACC CGG TGG CAC CGT TCC CGT CCG GCT GGT CGA CGA CGA CCT GGC AAA CTC GCT GGC
gly thr arg trp his arg ser arg pro ala gly arg arg arg pro gly lys leu ala gly
301/101                                331/111
CAA CGG CGG CCG GCT GGA CAT CCT CCT GTC GGC GGC CGA GTT CGC CAC CAA CCG GGA AGT
gln arg arg pro ala gly his pro pro val gly gly arg val arg his gln pro gly ser
361/121                                391/131
CGA CCC CGA CGG CGC CGT CGG CCG AGC GCT GTG CCT GGC CAT CGA CCC AGA TC
arg pro arg arg arg arg arg pro ser ala val pro gly his arg pro arg

```

SEQ ID N° 36 C

FIGURE 36C

Séquence codante Rv3909 prédite par Cole et al., 1998 (Nature 393 537-544)
contenant Seq 36A

```

1/1                                31/11
GTG ACC GCA CTG CAA CTC GGC TGG GCC GCT TTG GCG CGC GTC ACC TCA GCG ATC GGC GTC
met thr ala leu gln leu gly trp ala ala leu ala arg val thr ser ala ile gly val
61/21                                91/31
GTG GCC GGC CTC GGG ATG GCG CTC ACG GTA CCG TCG GCG GCA CCG CAC GCG CTC GCA GGC
val ala gly leu gly met ala leu thr val pro ser ala ala pro his ala leu ala gly
121/41                                151/51
GAG CCC AGC CCG ACG CCT TTT GTC CAG GTC CGC ATC GAT CAG GTG ACC CCG GAC GTG GTG
glu pro ser pro thr pro phe val gln val arg ile asp gln val thr pro asp val val
181/61                                211/71
ACC ACT TCC AGC GAA CCC CAT GTC ACC GTC AGC GGA ACG GTG ACC AAT ACC GGT GAC CGC
thr thr ser ser glu pro his val thr val ser gly thr val thr asn thr gly asp arg
241/81                                271/91
CCA GTC CGC GAT GTG ATG GTC CGG CTT GAG CAC GCC GCC GCG GTC ACG TCG TCA ACG GCG
pro val arg asp val met val arg leu glu his ala ala ala val thr ser ser thr ala
301/101                                331/111
TTA CGC ACC TCG CTC GAC GGC GGC ACC GAC CAG TAC CAG CCG GCC GCG GAC TTC CTC ACG
leu arg thr ser leu asp gly gly thr asp gln tyr gln pro ala ala asp phe leu thr

```

SEQ ID N° 36D

FIGURE 36D

FEUILLE DE REMPLACEMENT (REGLE 26)

111/185

361/121
 GTC GCC CCC GAA CTA GAC CGC GGG CAA GAG GCC GGC TTT ACC CTC TCG GCC CCG CTG CGC
 val ala pro glu leu asp arg gly gln glu ala gly phe thr leu ser ala pro leu arg
 421/141
 TCG CTG ACC AGG CCG TCG TTG GCC GTC AAC CAG CCC GGG ATC TAC CCG GTC CTG GTC AAC
 ser leu thr arg pro ser leu ala val asn gln pro gly ile tyr pro val leu val asn
 481/161
 GTC AAT GGG ACA CCC GAC TAC GGT GCG CCT GCG CGG CTC GAC AAT GCG CGG TTC CTG TTG
 val asn gly thr pro asp tyr gly ala pro ala arg leu asp asn ala arg phe leu leu
 541/181
 CCC GTG GTC GGA GTG CCA CCC GAC CAG GCC ACC GAC TTC GGC TCC GCT GTT GCA CCA GAA
 pro val val gly val pro pro asp gln ala thr asp phe gly ser ala val ala pro glu
 601/201
 ACG ACG GCG CCG GTC TGG ATC ACC ATG CTG TGG CCG CTG GCC GAC CGG CCC CGG TTG GCC
 thr thr ala pro val trp ile thr met leu trp pro leu ala asp arg pro arg leu ala
 661/221
 CCC GGG GCA CCC GGT GGC ACC GTT CCC GTC CCG CTG GTC GAC GAC GAC CTG GCA AAC TCG
 pro gly ala pro gly gly thr val pro val arg leu val asp asp asp leu ala asn ser
 721/241
 CTG GCC AAC GGC GGC CGG CTG GAC ATC CTC CTG TCG GCG GCC GAG TTC GCC ACC AAC CGG
 leu ala asn gly gly arg leu asp ile leu leu ser ala ala glu phe ala thr asn arg
 781/261
 GAA GTC GAC CCC GAC GGC GCC GTC GGC CGA GCG CTG TGC CTG GCC ATC GAC CCA GAT CTA
 glu val asp pro asp gly ala val gly arg ala leu cys leu ala ile asp pro asp leu
 841/281
 CTC ATC ACC GTC AAT GCG ATG ACC GGC GGC TAC GTC GTG TCC GAC TCG CCC GAC GGG GCC
 leu ile thr val asn ala met thr gly gly tyr val val ser asp ser pro asp gly ala
 901/301
 GCT CAA CTA CCG GGC ACC CCG ACC CAC CCG GGC ACC GGC CAG GCC GCC GCA TCC AGC TGG
 ala gln leu pro gly thr pro thr his pro gly thr gly gln ala ala ala ser ser trp
 961/321
 CTG GAT CGA TTG CCG ACG CTA GTC CAC CGG ACA TGC GTG ACG CCG CTG CCT TTT GCC CAA
 leu asp arg leu arg thr leu val his arg thr cys val thr pro leu pro phe ala gln
 1021/341
 GCC GAC CTG GAT GCT TTG CAG CGG GTT AAT GAT CCG AGG CTG AGC GCG ATC GCA ACC ATC
 ala asp leu asp ala leu gln arg val asn asp pro arg leu ser ala ile ala thr ile
 1081/361
 AGC CCC GCC GAC ATC GTC GAC CGC ATC CTG GAT GTC AGC TCC ACC CGC GGC GCA ACC GTG
 ser pro ala asp ile val asp arg ile leu asp val ser ser thr arg gly ala thr val
 1141/381
 CTG CCC GAC GGC CCG TTG ACC GGC CGG GCG ATC AAC TTG CTC AGC ACC CAC GGC AAC ACG
 leu pro asp gly pro leu thr gly arg ala ile asn leu leu ser thr his gly asn thr
 1201/401
 GTT GCC GTC GCG GCC GCC GAT TTT AGC CCC GAG GAA CAG CAG GGT TCG TCC CAG ATC GGC
 val ala val ala ala ala asp phe ser pro glu glu gln gln gly ser ser gln ile gly
 1261/421
 TCC GCG CTC TTA CCC GCT ACC GCG CCC CGG CCG TTG TCC CCG CGG GTG GTA GCG GCG CCG
 ser ala leu leu pro ala thr ala pro arg arg leu ser pro arg val val ala ala pro
 1321/441
 TTT GAT CCC GCG GTC GGG GCC GCG CTG GCC GCC GCG GGA ACA AAC CCG ACC GTT CCT ACC
 phe asp pro ala val gly ala ala leu ala ala ala gly thr asn pro thr val pro thr
 1381/461
 TAT CTA GAT CCC TCG TTG TTC GTT CGG ATC GCG CAT GAA TCG ATC ACC GCG CGC CGC CAG
 tyr leu asp pro ser leu phe val arg ile ala his glu ser ile thr ala arg arg gln

SEQ ID N° 36D(suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

112/185

1441/481

GAC GCC TTG GGC GCA ATG CTG TGG CGC AGC
 asp ala leu gly ala met leu trp arg ser
 1501/501

ATC CTG GTG CCG CCG GCG TCG TGG AGC CTG
 ile leu val pro pro ala ser trp ser leu
 1561/521

GCG CTG GCC ACC GCC ATC CGG TCT GGC CTG
 ala leu ala thr ala ile arg ser gly leu
 1621/541

GCT GAC GCC GCG GCC CGC ACC GAG CCA CCG
 ala asp ala ala ala arg thr glu pro pro
 1681/561

GGC CGG TTC AAT GAC GAC ATC ACC ACG CAG
 gly arg phe asn asp asp ile thr thr gln
 1741/581

CTG ACC TCG GCG TTG ACC ATC GAT GAC CGC
 leu thr ser ala leu thr ile asp asp arg
 1801/601

CCA CTA CGC GAG GAC ATG TTG CGC GCG CTG
 pro leu arg glu asp met leu arg ala leu
 1861/621

GGG CTG GCC CAG CAG CGG CTG GCC GTC GTT
 gly leu ala gln gln arg leu ala val val
 1921/641

GTG ACC ATC GTC AAC CCG GGC GGC TCC TAC
 val thr ile val asn pro gly gly ser tyr
 1981/661

TTG GCG CTG CAT AAT GGC CTC GCC GTG CCA
 leu ala leu his asn gly leu ala val pro
 2041/681

CCC GGG ATG ACG GTG GCC GAT GTC GGT CAG
 pro gly met thr val ala asp val gly gln
 2101/701

CGA GTA CCA ATC GAG GTG AAC TTC ACA CAG
 arg val pro ile glu val asn phe thr gln
 2161/721

CCC GAC GGC GTC GCG CTG GGT GAA CCG GTG
 pro asp gly val ala leu gly glu pro val
 2221/741

AAG GTG TTG TTC GCG ATC ACG CTA TCC GCT
 lys val leu phe ala ile thr leu ser ala
 2281/761

CGC CTT TGG CAC CGG TTC CGT GGC CAG CCT
 arg leu trp his arg phe arg gly gln pro
 2341/781

CCT ACC GGC AAA CAC GCC CCG CAG CGC CGT
 pro thr gly lys his ala pro gln arg arg
 2401/801

CGG GTA TGA
 arg val OPA

1471/491

TTG GAG CCG AAT GCC GCG CCC CGT ACC CAA
 leu glu pro asn ala ala pro arg thr gln
 1531/511

GCC AGC GAC GAC GCG CAG GTC ATC CTG ACC
 ala ser asp asp ala gln val ile leu thr
 1591/531

GCC GTG CCG CGA CCA CTA CCG GCG GTG ATC
 ala val pro arg pro leu pro ala val ile
 1651/551

GAA CCC CCG GGC GCT TAC AGC GCC GCT CGC
 glu pro pro gly ala tyr ser ala ala arg
 1711/571

ATC GGC GGG CAG GTT GCC CGG CTA TGG AAG
 ile gly gly gln val ala arg leu trp lys
 1771/591

ACC GGG CTG ACC GGC GTG CAG TAC ACC GCA
 thr gly leu thr gly val gln tyr thr ala
 1831/611

AGC CAA TCG CTA CCA CCC GAT ACC CGC AAC
 ser gln ser leu pro pro asp thr arg asn
 1891/631

GGA AAG ACG ATC GAC GAT CTT TTC GGC GCG
 gly lys thr ile asp asp leu phe gly ala
 1951/651

ACT CTG GCC ACC GAG CAC AGT CCG CTG CCG
 thr leu ala thr glu his ser pro leu pro
 2011/671

ATC CGG GTC CGG CTA CAG GTC GAT GCT CCG
 ile arg val arg leu gln val asp ala pro
 2071/691

ATC GAG CTA CCG CCC GGG TAC CTG CCG CTA
 ile glu leu pro pro gly tyr leu pro leu
 2131/711

CGG GTT GCC GTC GAC GTG TCG CTG CGG ACC
 arg val ala val asp val ser leu arg thr
 2191/731

CGG TTG TCG GTG CAC TCC AAC GCC TAC GGC
 arg leu ser val his ser asn ala tyr gly
 2251/751

GCG GCC GTG CTG GTA ACG CTG GCG GGC CGG
 ala ala val leu val thr leu ala gly arg
 2311/771

GAT CGC GCC GAC CTG GAT CGC CCC GAC CTG
 asp arg ala asp leu asp arg pro asp leu
 2371/791

GCC GTA GCC AGT CGG GAT GAC GAA AAG CAC
 ala val ala ser arg asp asp glu lys his

SEQ ID N° 36D (suite 2)

FIGURE 36D (suite 2)

FEUILLE DE REMPLACEMENT (REGLE 26)

113/185

ORF d'après Cole et al., 1998 (Nature 393 537-544) et contenant Rv 3909.

```

1/1                                31/11
TGA CTC AGC ACC GGG TCA GCA CAA CGG TCC CGG GCC GGG GCC GTG ACC GCA CTG CAA CTC
OPA leu ser thr gly ser ala gln arg ser arg ala gly ala val thr ala leu gln leu
61/21                                91/31
GGC TGG GCC GCT TTG GCG CGC GTC ACC TCA GCG ATC GGC GTC GTG GCC GGC CTC GGG ATG
gly trp ala ala leu ala arg val thr ser ala ile gly val val ala gly leu gly met
121/41                                151/51
GCG CTC ACG GTA CCG TCG GCG GCA CCG CAC GCG CTC GCA GGC GAG CCC AGC CCG ACG CCT
ala leu thr val pro ser ala ala pro his ala leu ala gly glu pro ser pro thr pro
181/61                                211/71
TTT GTC CAG GTC CGC ATC GAT CAG GTG ACC CCG GAC GTG GTG ACC ACT TCC AGC GAA CCC
phe val gln val arg ile asp gln val thr pro asp val val thr thr ser ser glu pro
241/81                                271/91
CAT GTC ACC GTC AGC GGA ACG GTG ACC AAT ACC GGT GAC CGC CCA GTC CGC GAT GTG ATG
his val thr val ser gly thr val thr asn thr gly asp arg pro val arg asp val met
301/101                                331/111
GTC CGG CTT GAG CAC GCC GCC GCG GTC ACG TCG TCA ACG GCG TTA CGC ACC TCG CTC GAC
val arg leu glu his ala ala ala val thr ser ser thr ala leu arg thr ser leu asp
361/121                                391/131
GGC GGC ACC GAC CAG TAC CAG CCG GCC GCG GAC TTC CTC ACG GTC GCC CCC GAA CTA GAC
gly gly thr asp gln tyr gln pro ala ala asp phe leu thr val ala pro glu leu asp
421/141                                451/151
CGC GGG CAA GAG GCC GGC TTT ACC CTC TCG GCC CCG CTG CGC TCG CTG ACC AGG CCG TCG
arg gly gln glu ala gly phe thr leu ser ala pro leu arg ser leu thr arg pro ser
481/161                                511/171
TTG GCC GTC AAC CAG CCC GGG ATC TAC CCG GTC CTG GTC AAC GTC AAT GGG ACA CCC GAC
leu ala val asn gln pro gly ile tyr pro val leu val asn val asn gly thr pro asp
541/181                                571/191
TAC GGT GCG CCT GCG CGG CTC GAC AAT GCG CGG TTC CTG TTG CCC GTG GTC GGA GTG CCA
tyr gly ala pro ala arg leu asp asn ala arg phe leu leu pro val val gly val pro
601/201                                631/211
CCC GAC CAG GCC ACC GAC TTC GGC TCC GCT GTT GCA CCA GAA ACG ACG GCG CCG GTC TGG
pro asp gln ala thr asp phe gly ser ala val ala pro glu thr thr ala pro val trp
661/221                                691/231
ATC ACC ATG CTG TGG CCG CTG GCC GAC CGG CCC CGG TTG GCC CCC GGG GCA CCC GGT GGC
ile thr met leu trp pro leu ala asp arg pro arg leu ala pro gly ala pro gly gly
721/241                                751/251
ACC GTT CCC GTC CGG CTG GTC GAC GAC GAC CTG GCA AAC TCG CTG GCC AAC GGC GGC CGG
thr val pro val arg leu val asp asp asp leu ala asn ser leu ala asn gly gly arg
781/261                                811/271
CTG GAC ATC CTC CTG TCG GCG GCC GAG TTC GCC ACC AAC CGG GAA GTC GAC CCC GAC GGC
leu asp ile leu leu ser ala ala glu phe ala thr asn arg glu val asp pro asp gly
841/281                                871/291
GCC GTC GGC CGA GCG CTG TGC CTG GCC ATC GAC CCA GAT CTA CTC ATC ACC GTC AAT GCG
ala val gly arg ala leu cys leu ala ile asp pro asp leu leu ile thr val asn ala
901/301                                931/311
ATG ACC GGC GGC TAC GTC GTG TCC GAC TCG CCC GAC GGG GCC GCT CAA CTA CCG GGC ACC
met thr gly gly tyr val val ser asp ser pro asp gly ala ala gln leu pro gly thr
961/321                                991/331
CCG ACC CAC CCG GGC ACC GGC CAG GCC GCC GCA TCC AGC TGG CTG GAT CGA TTG CGG ACG
pro thr his pro gly thr gly gln ala ala ala ser ser trp leu asp arg leu arg thr

```

SEQ ID N° 36F

FEUILLE DE REMPLACEMENT (REGLE 26)

114/185

1021/341	1051/351
CTA GTC CAC CGG ACA TGC GTG ACG CCG CTG	CCT TTT GCC CAA GCC GAC CTG GAT GCT TTG
leu val his arg thr cys val thr pro leu	pro phe ala gln ala asp leu asp ala leu
1081/361	1111/371
CAG CGG GTT AAT GAT CCG AGG CTG AGC GCG	ATC GCA ACC ATC AGC CCC GCC GAC ATC GTC
gln arg val asn asp pro arg leu ser ala	ile ala thr ile ser pro ala asp ile val
1141/381	1171/391
GAC CGC ATC CTG GAT GTC AGC TCC ACC CGC	GGC GCA ACC GTG CTG CCC GAC GGC CCG TTG
asp arg ile leu asp val ser ser thr arg	gly ala thr val leu pro asp gly pro leu
1201/401	1231/411
ACC GGC CGG GCG ATC AAC TTG CTC AGC ACC	CAC GGC AAC ACG GTT GCC GTC GCG GCC GCC
thr gly arg ala ile asn leu leu ser thr	his gly asn thr val ala val ala ala ala
1261/421	1291/431
GAT TTT AGC CCC GAG GAA CAG CAG GGT TCG	TCC CAG ATC GGC TCC GCG CTC TTA CCC GCT
asp phe ser pro glu glu gln gln gly ser	ser gln ile gly ser ala leu leu pro ala
1321/441	1351/451
ACC GCG CCC CGG CGG TTG TCC CCG CGG GTG	GTA GCG GCG CCG TTT GAT CCC GCG GTC GGG
thr ala pro arg arg leu ser pro arg val	val ala ala pro phe asp pro ala val gly
1381/461	1411/471
GCC GCG CTG GCC GCC GCG GGA ACA AAC CCG	ACC GTT CCT ACC TAT CTA GAT CCC TCG TTG
ala ala leu ala ala ala gly thr asn pro	thr val pro thr tyr leu asp pro ser leu
1441/481	1471/491
TTC GTT CGG ATC GCG CAT GAA TCG ATC ACC	GCG CGC CGC CAG GAC GCC TTG GGC GCA ATG
phe val arg ile ala his glu ser ile thr	ala arg arg gln asp ala leu gly ala met
1501/501	1531/511
CTG TGG CGC AGC TTG GAG CCG AAT GCC GCG	CCC CGT ACC CAA ATC CTG GTG CCG CCG GCG
leu trp arg ser leu glu pro asn ala ala	pro arg thr gln ile leu val pro pro ala
1561/521	1591/531
TCG TGG AGC CTG GCC AGC GAC GAC GCG CAG	GTC ATC CTG ACC GCG CTG GCC ACC GCC ATC
ser trp ser leu ala ser asp asp ala gln	val ile leu thr ala leu ala thr ala ile
1621/541	1651/551
CGG TCT GGC CTG GCC GTG CCG CGA CCA CTA	CCG GCG GTG ATC GCT GAC GCC GCG GCC CGC
arg ser gly leu ala val pro arg pro leu	pro ala val ile ala asp ala ala ala arg
1681/561	1711/571
ACC GAG CCA CCG GAA CCC CCG GGC GCT TAC	AGC GCC GCT CGC GGC CGG TTC AAT GAC GAC
thr glu pro pro glu pro pro gly ala tyr	ser ala ala arg gly arg phe asn asp asp
1741/581	1771/591
ATC ACC ACG CAG ATC GGC GGG CAG GTT GCC	CGG CTA TGG AAG CTG ACC TCG GCG TTG ACC
ile thr thr gln ile gly gly gln val ala	arg leu trp lys leu thr ser ala leu thr
1801/601	1831/611
ATC GAT GAC CGC ACC GGG CTG ACC GGC GTG	CAG TAC ACC GCA CCA CTA CGC GAG GAC ATG
ile asp asp arg thr gly leu thr gly val	gln tyr thr ala pro leu arg glu asp met
1861/621	1891/631
TTG CGC GCG CTG AGC CAA TCG CTA CCA CCC	GAT ACC CGC AAC GGG CTG GCC CAG CAG CGG
leu arg ala leu ser gln ser leu pro pro	asp thr arg asn gly leu ala gln gln arg
1921/641	1951/651
CTG GCC GTC GTT GGA AAG ACG ATC GAC GAT	CTT TTC GGC GCG GTG ACC ATC GTC AAC CCG
leu ala val val gly lys thr ile asp asp	leu phe gly ala val thr ile val asn pro
1981/661	2011/671
GGC GGC TCC TAC ACT CTG GCC ACC GAG CAC	AGT CCG CTG CCG TTG GCG CTG CAT AAT GGC
gly gly ser tyr thr leu ala thr glu his	ser pro leu pro leu ala leu his asn gly

SEQ ID 36F (suite 1)

FIGURE 36F (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

115/185

2041/681	2071/691
CTC GCC GTG CCA ATC CGG GTC CGG CTA CAG	GTC GAT GCT CCG CCC GGG ATG ACG GTG GCC
leu ala val pro ile arg val arg leu gln	val asp ala pro pro gly met thr val ala
2101/701	2131/711
GAT GTC GGT CAG ATC GAG CTA CCG CCC GGG	TAC CTG CCG CTA CGA GTA CCA ATC GAG GTG
asp val gly gln ile glu leu pro pro gly	tyr leu pro leu arg val pro ile glu val
2161/721	2191/731
AAC TTC ACA CAG CGG GTT GCC GTC GAC GTG	TCG CTG CCG ACC CCC GAC GGC GTC GCG CTG
asn phe thr gln arg val ala val asp val	ser leu arg thr pro asp gly val ala leu
2221/741	2251/751
GGT GAA CCG GTG CGG TTG TCG GTG CAC TCC	AAC GCC TAC GGC AAG GTG TTG TTC GCG ATC
gly glu pro val arg leu ser val his ser	asn ala tyr gly lys val leu phe ala ile
2281/761	2311/771
ACG CTA TCC GCT GCG GCC GTG CTG GTA ACG	CTG GCG GGC CCG CGC CTT TGG CAC CGG TTC
thr leu ser ala ala ala val leu val thr	leu ala gly arg arg leu trp his arg phe
2341/781	2371/791
CGT GGC CAG CCT GAT CGC GCC GAC CTG GAT	CGC CCC GAC CTG CCT ACC GGC AAA CAC GCC
arg gly gln pro asp arg ala asp leu asp	arg pro asp leu pro thr gly lys his ala
2401/801	2431/811
CCG CAG CGC CGT GCC GTA GCC AGT CGG GAT	GAC GAA AAG CAC CGG GTA TGA
pro gln arg arg ala val ala ser arg asp	asp glu lys his arg val OPA

SEQ ID 36F (suite 2)

FIGURE 36F (suite 2)

1/1	31/11
ATC CGC GCG TTG GCG TCG CAT CCG AAC ATC	GTC GGA GTC AAG GAC GCC AAA GCC GAC CTG
ile arg ala leu ala ser his pro asn ile	val gly val lys asp ala lys ala asp leu
61/21	91/31
CAC AGC GGC GCC CAA ATC ATG GCC GAC ACC	GGA CTG GCC TAC TAT TCC GGC GAC GAC GCG
his ser gly ala gln ile met ala asp thr	gly leu ala tyr tyr ser gly asp asp ala
121/41	151/51
CTC AAC CTG CCC TGG CTG GCC ATG GGC GCC	ACG GGC TTC ATC AGC GTG ATT GCC CAC CTG
leu asn leu pro trp leu ala met gly ala	thr gly phe ile ser val ile ala his leu
181/61	211/71
GCA GCC GGG CAG CTT CGA GAG TTG TTG TCC	GCC TTC GGT TCT GGG GAT ATC GCC ACC GCC
ala ala gly gln leu arg glu leu leu ser	ala phe gly ser gly asp ile ala thr ala
241/81	
CGC AAG ATC	
arg lys ile	

SEQ ID N° 37A

FIGURE 37A

116/185

1/1 31/11
 GAT CCG CGC GTT GGC GTC GCA TCC GAA CAT CGT CGG AGT CAA GGA CGC CAA AGC CGA CCT
 asp pro arg val gly val ala ser glu his arg arg ser gln gly arg gln ser arg pro
 61/21 91/31
 GCA CAG CGG CGC CCA AAT CAT GGC CGA CAC CGG ACT GGC CTA CTA TTC CGG CGA CGA CGC
 ala gln arg arg pro asn his gly arg his arg thr gly leu leu phe arg arg arg arg
 121/41 151/51
 GCT CAA CCT GCC CTG GCT GGC CAT GGG CGC CAC GGG CTT CAT CAG CGT GAT TGC CCA CCT
 ala gln pro ala leu ala gly his gly arg his gly leu his gln arg asp cys pro pro
 181/61 211/71
 GGC AGC CGG GCA GCT TCG AGA GTT GTT GTC CGC CTT CGG TTC TGG GGA TAT CGC CAC CGC
 gly ser arg ala ala ser arg val val val arg leu arg phe trp gly tyr arg his arg
 241/81
 CCG CAA GAT C
 pro gln asp

SEQ ID N° 37B

FIGURE 37B

1/1 31/11
 TCC GCG CGT TGG CGT CGC ATC CGA ACA TCG TCG GAG TCA AGG ACG CCA AAG CCG ACC TGC
 ser ala arg trp arg arg ile arg thr ser ser glu ser arg thr pro lys pro thr cys
 61/21 91/31
 ACA GCG GCG CCC AAA TCA TGG CCG ACA CCG GAC TGG CCT ACT ATT CCG GCG ACG ACG CGC
 thr ala ala pro lys ser trp pro thr pro asp trp pro thr ile pro ala thr thr arg
 121/41 151/51
 TCA ACC TGC CCT GGC TGG CCA TGG GCG CCA CGG GCT TCA TCA GCG TGA TTG CCC ACC TGG
 ser thr cys pro gly trp pro trp ala pro arg ala ser ser ala OPA leu pro thr trp
 181/61 211/71
 CAG CCG GGC AGC TTC GAG AGT TGT TGT CCG CCT TCG GTT CTG GGG ATA TCG CCA CCG CCC
 gln pro gly ser phe glu ser cys cys pro pro ser val leu gly ile ser pro pro pro
 241/81
 GCA AGA TC
 ala arg

SEQ ID N° 37C

FIGURE 37C

117/185

Séquence codante Rv2753c prédite par Cole et al., 1998 (Nature 393 537-544)
contenant Seq 37A

```

1/1                                31/11
GTG ACC ACC GTC GGA TTC GAC GTC GCA GCG CGC CTA GGA ACC CTG CTG ACC GCG ATG GTG
val thr thr val gly phe asp val ala ala arg leu gly thr leu leu thr ala met val
61/21                                91/31
ACA CCG TTT AGC GGC GAT GGC TCC CTG GAC ACC GCC ACC GCG GCG CGG CTG GCC AAC CAC
thr pro phe ser gly asp gly ser leu asp thr ala thr ala ala arg leu ala asn his
121/41                                151/51
CTG GTC GAT CAG GGG TGC GAC GGT CTG GTG GTC TCG GGC ACC ACC GGC GAG TCG CCG ACC
leu val asp gln gly cys asp gly leu val val ser gly thr thr gly glu ser pro thr
181/61                                211/71
ACC ACC GAC GGG GAG AAA ATC GAG CTG CTG CGG GCC GTC TTG GAA GCG GTG GGG GAC CGG
thr thr asp gly glu lys ile glu leu leu arg ala val leu glu ala val gly asp arg
241/81                                271/91
GCC CGT GTT ATC GCC GGT GCC GGC ACC TAT GAC ACC GCG CAC AGC ATC CGG CTG GCC AAG
ala arg val ile ala gly ala gly thr tyr asp thr ala his ser ile arg leu ala lys
301/101                                331/111
GCT TGT GCG GCC GAG GGT GCG CAC GGG CTG CTG GTG GTC ACG CCC TAC TAT TCC AAG CCG
ala cys ala ala glu gly ala his gly leu leu val thr pro tyr tyr ser lys pro
361/121                                391/131
CCG CAG CGG GGG CTG CAA GCC CAT TTC ACC GCC GTC GCC GAC GCG ACC GAG CTG CCG ATG
pro gln arg gly leu gln ala his phe thr ala val ala asp ala thr glu leu pro met
421/141                                451/151
CTG CTC TAT GAC ATC CCG GGG CGG TCG GCG GTG CCG ATC GAG CCC GAC ACG ATC CGC GCG
leu leu tyr asp ile pro gly arg ser ala val pro ile glu pro asp thr ile arg ala
481/161                                511/171
TTG GCG TCG CAT CCG AAC ATC GTC GGA GTC AAG GAC GCC AAA GCC GAC CTG CAC AGC GGC
leu ala ser his pro asn ile val gly val lys asp ala lys ala asp leu his ser gly
541/181                                571/191
GCC CAA ATC ATG GCC GAC ACC GGA CTG GCC TAC TAT TCC GGC GAC GAC GCG CTC AAC CTG
ala gln ile met ala asp thr gly leu ala tyr tyr ser gly asp asp ala leu asn leu
601/201                                631/211
CCC TGG CTG GCC ATG GGC GCC ACG GGC TTC ATC AGC GTG ATT GCC CAC CTG GCA GCC GGG
pro trp leu ala met gly ala thr gly phe ile ser val ile ala his leu ala ala gly
661/221                                691/231
CAG CTT CGA GAG TTG TTG TCC GCC TTC GGT TCT GGG GAT ATC GCC ACC GCC CGC AAG ATC
gln leu arg glu leu leu ser ala phe gly ser gly asp ile ala thr ala arg lys ile
721/241                                751/251
AAC ATT GCG GTC GCC CCG CTG TGC AAC GCG ATG AGC CGC CTG GGT GGG GTG ACG TTG TCC
asn ile ala val ala pro leu cys asn ala met ser arg leu gly gly val thr leu ser
781/261                                811/271
AAG GCG GGC TTG CGG CTG CAG GGC ATC GAC GTC GGT GAT CCC CGG CTG CCC CAG GTG GCC
lys ala gly leu arg leu gln gly ile asp val gly asp pro arg leu pro gln val ala
841/281                                871/291
GCG ACA CCG GAG CAG ATC GAC GCG TTG GCC GCC GAC ATG CGC GCG GCC TCG GTG CTT CGG
ala thr pro glu gln ile asp ala leu ala ala asp met arg ala ala ser val leu arg

```

901/301
TGA
OPA

SEQ ID N° 37D

FIGURE 37D

FEUILLE DE REMPLACEMENT (REGLE 26)

118/185

ORF d'après Cole et al., 1998 (Nature 393 537-544) contenant Rv2753c

```

1/1                                31/11
TAA GGT GAG CGC CGT GGC CGA GAC CGC GCC GCT GCG CGT GCA ACT GAT CGC CAA GAC CGA
OCH gly glu arg arg gly arg asp arg ala ala ala arg ala thr asp arg gln asp arg
61/21                                91/31
CTT CTT GGC CCC ACC CGA CGT GCC CTG GAC CAC CGA CGC CGA CGG CGG ACC CGC GCT GGT
leu leu gly pro thr arg arg ala leu asp his arg arg arg arg arg thr arg ala gly
121/41                                151/51
CGA GTT CGC CGG CCG GGC CTG CTA TCA GAG CTG GTC CAA GCC CAA TCC CAA GAC CGC CAC
arg val arg arg pro gly leu leu ser glu leu val gln ala gln ser gln asp arg his
181/61                                211/71
CAA CGC CGG CTA CCT CCG GCA CAT CAT CGA CGT CGG ACA TTT CTC GGT GCT AGA GCA TGC
gln arg arg leu pro pro ala his his arg arg arg thr phe leu gly ala arg ala cys
241/81                                271/91
CAG CGT GTC GTT CTA CAT CAC CGG GAT CTC GCG ATC GTG CAC CCA CGA GCT GAT CCG CCA
gln arg val val leu his his arg asp leu ala ile val his pro arg ala asp pro pro
301/101                                331/111
CCG GCA TTT CTC CTA CTC GCA GCT CTC CCA GCG CTA CGT ACC CGA GAA GGA CTC GCG GGT
pro ala phe leu leu leu ala ala leu pro ala leu arg thr arg glu gly leu ala gly
361/121                                391/131
CGT CGT GCC GCC CGG CAT GGA GGA CGA CGC CGA CCT GCG CCA CAT CCT GAC CGA GGC CGC
arg arg ala ala arg his gly gly arg arg arg pro ala pro his pro asp arg gly arg
421/141                                451/151
CGA CGC CGC CCG CGC CAC CTA CAG CGA GCT GCT GGC CAA GCT GGA AGC CAA GTT CGC CGA
arg arg arg pro arg his leu gln arg ala ala gly gln ala gly ser gln val arg arg
481/161                                511/171
CCA ACC CAA CGC GAT CCT GCG CCG CAA GCA GGC CCG CCA AGC CGC CCG CGC GGT GCT GCC
pro thr gln arg asp pro ala pro gln ala gly pro pro ser arg pro arg gly ala ala
541/181                                571/191
CAA CGC CAC CGA AAC CCG CAT CGT GGT GAC CGG CAA CTA CCG GGC CTG GCG GCA CTT CAT
gln arg his arg asn pro his arg gly asp arg gln leu pro gly leu ala ala leu his
601/201                                631/211
CGC AAT GCG GGC CAG CGA GCA CGC CGA CGT GGA AAT CCG GCG ACT GGC CAT CGA ATG CCT
arg asn ala gly gln arg ala arg arg arg gly asn pro ala thr gly his arg met pro
661/221                                691/231
GCG CCA GCT CGC CGC CGT GGC CCC CGC GGT GTT CGC CGA CTT CGA GGT GAC CAC CCT GGC
ala pro ala arg arg arg gly pro arg gly val arg arg leu arg gly asp his pro gly
721/241                                751/251
CGA CGG CAC CGA GGT GGC GAC CAG CCC GTT GGC GAC CGA AGC CTG AGG CGG CGT GTC GCT
arg arg his arg gly gly asp gln pro val gly asp arg ser leu arg arg arg val ala
781/261                                811/271
GGA CAA ACA CGC GCG CTC GCG GCC GGG ATA AAG CGC CAG GTA ACC TTG GGA GCC GTG ACC
gly gln thr arg ala leu ala ala gly ile lys arg gln val thr leu gly ala val thr
841/281                                871/291
ACC GTC GGA TTC GAC GTC GCA GCG CGC CTA GGA ACC CTG CTG ACC GCG ATG GTG ACA CCG
thr val gly phe asp val ala ala arg leu gly thr leu leu thr ala met val thr pro
901/301                                931/311
TTT AGC GGC GAT GGC TCC CTG GAC ACC GCC ACC GCG GCG CGG CTG GCC AAC CAC CTG GTC
phe ser gly asp gly ser leu asp thr ala thr ala ala arg leu ala asn his leu val
961/321                                991/331
GAT CAG GGG TGC GAC GGT CTG GTG GTC TCG GGC ACC ACC GGC GAG TCG CCG ACC ACC ACC
asp gln gly cys asp gly leu val val ser gly thr thr gly glu ser pro thr thr thr

```

SEQ ID N° 37F

FIGURE 37F
FEUILLE DE REMPLACEMENT (REGLE 26)

119/185

1021/341	1051/351
GAC GGG GAG AAA ATC GAG CTG CTG CGG GCC GTC TTG GAA GCG GTG GGG GAC CGG GCC CGT	asp gly glu lys ile glu leu leu arg ala val leu glu ala val gly asp arg ala arg
1081/361	1111/371
GTT ATC GCC GGT GCC GGC ACC TAT GAC ACC GCG CAC AGC ATC CGG CTG GCC AAG GCT TGT	val ile ala gly ala gly thr tyr asp thr ala his ser ile arg leu ala lys ala cys
1141/381	1171/391
GCG GCC GAG GGT GCG CAC GGG CTG CTG GTG GTC ACG CCC TAC TAT TCC AAG CCG CCG CAG	ala ala glu gly ala his gly leu leu val val thr pro tyr tyr ser lys pro pro gln
1201/401	1231/411
CGG GGG CTG CAA GCC CAT TTC ACC GCC GTC GCC GAC GCG ACC GAG CTG CCG ATG CTG CTC	arg gly leu gln ala his phe thr ala val ala asp ala thr glu leu pro met leu leu
1261/421	1291/431
TAT GAC ATC CCG GGG CGG TCG GCG GTG CCG ATC GAG CCC GAC ACG ATC CGC GCG TTG GCG	tyr asp ile pro gly arg ser ala val pro ile glu pro asp thr ile arg ala leu ala
1321/441	1351/451
TCG CAT CCG AAC ATC GTC GGA GTC AAG GAC GCC AAA GCC GAC CTG CAC AGC GGC GCC CAA	ser his pro asn ile val gly val lys asp ala lys ala asp leu his ser gly ala gln
1381/461	1411/471
ATC ATG GCC GAC ACC GGA CTG GCC TAC TAT TCC GGC GAC GAC GCG CTC AAC CTG CCC TGG	ile met ala asp thr gly leu ala tyr tyr ser gly asp asp ala leu asn leu pro trp
1441/481	1471/491
CTG GCC ATG GGC GCC ACG GGC TTC ATC AGC GTG ATT GCC CAC CTG GCA GCC GGG CAG CTT	leu ala met gly ala thr gly phe ile ser val ile ala his leu ala ala gly gln leu
1501/501	1531/511
CGA GAG TTG TTG TCC GCC TTC GGT TCT GGG GAT ATC GCC ACC GCC CGC AAG ATC AAC ATT	arg glu leu leu ser ala phe gly ser gly asp ile ala thr ala arg lys ile asn ile
1561/521	1591/531
GCG GTC GCC CCG CTG TGC AAC GCG ATG AGC CGC CTG GGT GGG GTG ACG TTG TCC AAG GCG	ala val ala pro leu cys asn ala met ser arg leu gly gly val thr leu ser lys ala
1621/541	1651/551
GGC TTG CGG CTG CAG GGC ATC GAC GTC GGT GAT CCC CGG CTG CCC CAG GTG GCC GCG ACA	gly leu arg leu gln gly ile asp val gly asp pro arg leu pro gln val ala ala thr
1681/561	1711/571
CCG GAG CAG ATC GAC GCG TTG GCC GCC GAC ATG CGC GCG GCC TCG GTG CTT CGG TGA	pro glu gln ile asp ala leu ala ala asp met arg ala ala ser val leu arg OPA

SEQ ID N° 37F (suite 1)

FIGURE 37F (suite 1)

120/185

1/1 31/11
 GCG GTG AAC TGG TGG GCC CGG ATG GTT CAA GTA CGC CGT CGC AAA CTC GAG CAC AAC AGG
 ala val asn trp trp ala arg met val gln val arg arg arg lys leu glu his asn arg
 61/21 91/31
 AGA CGA CGG ATG GAA GGA GAT GCT GGC GCC GGC CAG CTG AAC CCT GCC GAT GCG AAT AAG
 arg arg arg met glu gly asp ala gly ala gly gln leu asn pro ala asp ala asn lys
 121/41 151/51
 TCG TCG TCT ACG GAG GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GCC GAC CAG
 ser ser ser thr glu val lys ala ala asp ser ala glu ser asp ala gly ala asp gln
 181/61 211/71
 ACT GGC CCG CAG GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GAG CTC GGC GAG
 thr gly pro gln val lys ala ala asp ser ala glu ser asp ala gly glu leu gly glu
 241/81 271/91
 GAC GCG TGC CCA GAA CAG GCC CTC GTC GAG CGG CGC CCG TCG CGG TTG CGG CGA GGC TGG
 asp ala cys pro glu gln ala leu val glu arg arg pro ser arg leu arg arg gly trp
 301/101 331/111
 CTT GTT GGC ATT GCG GCG ACG CTG CTC GCG TTG GCC GGT GGC CTT GGC GCA GCG GGT TAT
 leu val gly ile ala ala thr leu leu ala leu ala gly gly leu gly ala ala gly tyr
 361/121 391/131
 TTT GCG TTG CGC TCA CAC CAG GAA AGC CAA TCA ATC GCG CGC GAG GAC CTT GCG GCC ATT
 phe ala leu arg ser his gln glu ser gln ser ile ala arg glu asp leu ala ala ile
 421/141 451/151
 GAG GCC GCT AAG GAT TGC GTT GCG GCC ACG CAG GCA CCC GAT GCT GGG GCG ATG TCG GCT
 glu ala ala lys asp cys val ala ala thr gln ala pro asp ala gly ala met ser ala
 481/161
 AGC ATG CAG AAG ATC
 ser met gln lys ile

SEQ ID N° 38A

FIGURE 38A

1/1 31/11
 CAG CGG TGA ACT GGT GGG CCC GGA TGG TTC AAG TAC GCC GTC GCA AAC TCG AGC ACA ACA
 gln arg OPA thr gly gly pro gly trp phe lys tyr ala val ala asn ser ser thr thr
 61/21 91/31
 GGA GAC GAC GGA TGG AAG GAG ATG CTG GCG CCG GCC AGC TGA ACC CTG CCG ATG CGA ATA
 gly asp asp gly trp lys glu met leu ala pro ala ser OPA thr leu pro met arg ile
 121/41 151/51
 AGT CGT CGT CTA CGG AGG TGA AGG CGG CGG ATT CGG CGG AAT CTG ACG CCG GAG CCG ACC
 ser arg arg leu arg arg OPA arg arg arg ile arg arg asn leu thr pro glu pro thr
 181/61 211/71
 AGA CTG GCC CGC AGG TGA AGG CGG CGG ATT CGG CGG AAT CTG ACG CCG GAG AGC TCG GCG
 arg leu ala arg arg OPA arg arg arg ile arg arg asn leu thr pro glu ser ser ala
 241/81 271/91
 AGG ACG CGT GCC CAG AAC AGG CCC TCG TCG AGC GGC GCC CGT CGC GGT TGC GGC GAG GCT
 arg thr arg ala gln asn arg pro ser ser ser gly ala arg arg gly cys gly glu ala
 301/101 331/111
 GGC TTG TTG GCA TTG CGG CGA CGC TGC TCG CGT TGG CCG GTG GCC TTG GCG CAG CGG GTT
 gly leu leu ala leu arg arg arg cys ser arg trp pro val ala leu ala gln arg val
 361/121 391/131
 ATT TTG CGT TGC GCT CAC ACC AGG AAA GCC AAT CAA TCG CGC GCG AGG ACC TTG CGG CCA
 ile leu arg cys ala his thr arg lys ala asn gln ser arg ala arg thr leu arg pro
 421/141 451/151
 TTG AGG CCG CTA AGG ATT GCG TTG CGG CCA CGC AGG CAC CCG ATG CTG GGG CGA TGT CGG
 leu arg pro leu arg ile ala leu arg pro arg arg his pro met leu gly arg cys arg
 481/161
 CTA GCA TGC AGA AGA TC
 leu ala cys arg arg

SEQ ID N° 38B

FEUILLE DE REMPLACEMENT (REGLE 26)

121/185

```

1/1
AGC GGT GAA CTG GTG GGC CCG GAT GGT TCA AGT ACG CCG TCG CAA ACT CGA GCA CAA CAG
ser gly glu leu val gly pro asp gly ser ser thr pro ser gln thr arg ala gln gln
61/21
GAG ACG ACG GAT GGA AGG AGA TGC TGG CGC CGG CCA GCT GAA CCC TGC CGA TGC GAA TAA
glu thr thr asp gly arg arg cys trp arg arg pro ala glu pro cys arg cys glu OCH
121/41
GTC GTC GTC TAC GGA GGT GAA GGC GGC GGA TTC GGC GGA ATC TGA CGC CGG AGC CGA CCA
val val val tyr gly gly glu gly gly gly phe gly gly ile OPA arg arg ser arg pro
181/61
GAC TGG CCC GCA GGT GAA GGC GGC GGA TTC GGC GGA ATC TGA CGC CGG AGA GCT CGG CGA
asp trp pro ala gly glu gly gly gly phe gly gly ile OPA arg arg arg ala arg arg
241/81
GGA CGC GTG CCC AGA ACA GGC CCT CGT CGA GCG GCG CCC GTC GCG GTT GCG GCG AGG CTG
gly arg val pro arg thr gly pro arg arg ala ala pro val ala val ala ala arg leu
301/101
GCT TGT TGG CAT TGC GGC GAC GCT GCT CGC GTT GGC CGG TGG CCT TGG CGC AGC GGG TTA
ala cys trp his cys gly asp ala ala arg val gly arg trp pro trp arg ser gly leu
361/121
TTT TGC GTT GCG CTC ACA CCA GGA AAG CCA ATC AAT CGC GCG CGA GGA CCT TGC GGC CAT
phe cys val ala leu thr pro gly lys pro ile asn arg ala arg gly pro cys gly his
421/141
TGA GGC CGC TAA GGA TTG CGT TGC GGC CAC GCA GGC ACC CGA TGC TGG GGC GAT GTC GGC
OPA gly arg OCH gly leu arg cys gly his ala gly thr arg cys trp gly asp val gly
481/161
TAG CAT GCA GAA GAT C
AMB his ala glu asp

```

SEQ ID N° 38C

FIGURE 38C

122/185

Séquence Rv0175 prédite par Cole et al., 1998 (Nature 393 537-544) et contenant seq38A

1/1	31/11
GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC	GCC GGA GCC GAC CAG ACT GGC CCG CAG GTG
val lys ala ala asp ser ala glu ser asp	ala gly ala asp gln thr gly pro gln val
61/21	91/31
AAG GCG GCG GAT TCG GCG GAA TCT GAC	GCC GGA GAG CTC GGC GAG GAC GCG TGC CCA GAA
lys ala ala asp ser ala glu ser asp ala	gly glu leu gly glu asp ala cys pro glu
121/41	151/51
CAG GCC CTC GTC GAG CGG CGC CCG TCG CGG	TTG CGG CGA GGC TGG CTT GTT GGC ATT GCG
gln ala leu val glu arg arg pro ser arg	leu arg arg gly trp leu val gly ile ala
181/61	211/71
GCG ACG CTG CTC GCG TTG GCC GGT GGC CTT	GGC GCA GCG GGT TAT TTT GCG TTG CGC TCA
ala thr leu leu ala leu ala gly gly leu	gly ala ala gly tyr phe ala leu arg ser
241/81	271/91
CAC CAG GAA AGC CAA TCA ATC GCG CGC GAG	GAC CTT GCG GCC ATT GAG GCC GCT AAG GAT
his gln glu ser gln ser ile ala arg glu	asp leu ala ala ile glu ala ala lys asp
301/101	331/111
TGC GTT GCG GCC ACG CAG GCA CCC GAT GCT	GGG GCG ATG TCG GCT AGC ATG CAG AAG ATC
cys val ala ala thr gln ala pro asp ala	gly ala met ser ala ser met gln lys ile
361/121	391/131
ATC GAG TGT GGC ACC GGT GAT TTC GGT GCC	CAG GCG TCG TTG TAC ACC AGC ATG CTC GTC
ile glu cys gly thr gly asp phe gly ala	gln ala ser leu tyr thr ser met leu val
421/141	451/151
GAG GCG TAT CAA GCG GCC AGC GTC CAC GTG	CAA GTG ACC GAT ATG CGC GCG GCG GTC GAG
glu ala tyr gln ala ala ser val his val	gln val thr asp met arg ala ala val glu
481/161	511/171
CGC AAC AAC AAT GAC GGG TCG GTC GAT GTT	CTG GTG GCG CTC CGG GTC AAG GTG TCC AAC
arg asn asn asn asp gly ser val asp val	leu val ala leu arg val lys val ser asn
541/181	571/191
ACC GAC TCG GAT GCC CAT GAA GTC GGC TAC	CGT CTT CGG GTC CGG ATG GCA CTG GAT GAG
thr asp ser asp ala his glu val gly tyr	arg leu arg val arg met ala leu asp glu
601/201	631/211
GGC CGC TAT AAG ATC GCC AAA CTC GAC CAG	GTG ACG AAG TGA
gly arg tyr lys ile ala lys leu asp gln	val thr lys OPA

SEQ ID N° 38D

FIGURE 38D

123/185

ORF d'après Cole et al., 1998 (Nature 393 537-544) Contenant Rv0175

1/1
 TGA ACT GGT GGG GCC GGA TGG TGT CAA GTA CGC CGT CGC AAA CTC GAG CAC AAC AGG AGA
 OPA thr gly gly ala gly trp cys gln val arg arg arg lys leu glu his asn arg arg
 61/21
 CGA CGG ATG GAA GGA GAT GCT GGC GCC GGC CAG CTG AAC CCT GCC GAT GCG AAT AAG TCG
 arg arg met glu gly asp ala gly ala gly gln leu asn pro ala asp ala asn lys ser
 121/41
 TCG TCT ACG GAG GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GCC GAC CAG ACT
 ser ser thr glu val lys ala ala asp ser ala glu ser asp ala gly ala asp gln thr
 181/61
 GGC CCG CAG GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GAG CTC GGC GAG GAC
 gly pro gln val lys ala ala asp ser ala glu ser asp ala gly glu leu gly glu asp
 241/81
 GCG TGC CCA GAA CAG GCC CTC GTC GAG CGG CGC CCG TCG CGG TTG CGG CGA GGC TGG CTT
 ala cys pro glu gln ala leu val glu arg arg pro ser arg leu arg arg gly trp leu
 301/101
 GTT GGC ATT GCG GCG ACG CTG CTC GCG TTG GCC GGT GGC CTT GGC GCA GCG GGT TAT TTT
 val gly ile ala ala thr leu leu ala leu ala gly gly leu gly ala ala gly tyr phe
 361/121
 GCG TTG CGC TCA CAC CAG GAA AGC CAA TCA ATC GCG CGC GAG GAC CTT GCG GCC ATT GAG
 ala leu arg ser his gln glu ser gln ser ile ala arg glu asp leu ala ala ile glu
 421/141
 GCC GCT AAG GAT TGC GTT GCG GCC ACG CAG GCA CCC GAT GCT GGG GCG ATG TCG GCT AGC
 ala ala lys asp cys val ala ala thr gln ala pro asp ala gly ala met ser ala ser
 481/161
 ATG CAG AAG ATC ATC GAG TGT GGC ACC GGT GAT TTC GGT GCC CAG GCG TCG TTG TAC ACC
 met gln lys ile ile glu cys gly thr gly asp phe gly ala gln ala ser leu tyr thr
 541/181
 AGC ATG CTC GTC GAG GCG TAT CAA GCG GCC AGC GTC CAC GTG CAA GTG ACC GAT ATG CGC
 ser met leu val glu ala tyr gln ala ala ser val his val gln val thr asp met arg
 601/201
 GCG GCG GTC GAG CGC AAC AAC AAT GAC GGG TCG GTC GAT GTT CTG GTG GCG CTC CGG GTC
 ala ala val glu arg asn asn asn asp gly ser val asp val leu val ala leu arg val
 661/221
 AAG GTG TCC AAC ACC GAC TCG GAT GCC CAT GAA GTC GGC TAC CGT CTT CGG GTC CGG ATG
 lys val ser asn thr asp ser asp ala his glu val gly tyr arg leu arg val arg met
 721/241
 GCA CTG GAT GAG GGC CGC TAT AAG ATC GCC AAA CTC GAC CAG GTG ACG AAG TGA
 ala leu asp glu gly arg tyr lys ile ala lys leu asp gln val thr lys OPA

SEQ ID N° 38F

FIGURE 38F

124/185

1/1 31/11
 ACA CCT CCC CCC CCG CCG CCG CTG CCG CCG GTT CCC TTT CCC AAG GAA TGT CCG GCG CCG
 thr pro pro pro pro pro pro leu pro pro val pro phe pro lys glu cys pro ala pro
 61/21 91/31
 GGC GTG ATG CAA GGC TGC CTT GAG AGC ACC AGC GGC TTG ATC ATG GGC ATC GAC AGC AAG
 gly val met gln gly cys leu glu ser thr ser gly leu ile met gly ile asp ser lys
 121/41 151/51
 ACC GCA CTG GTC GCC GAG CGC ATC ACC GGT GCC GTC GAG GAG ATC
 thr ala leu val ala glu arg ile thr gly ala val glu glu ile

SEQ ID N° 39A

FIGURE 39A

1/1 31/11
 CAC CTC CCC CCC CGC CGC CGC TGC CGC CGG TTC CCT TTC CCA AGG AAT GTC CGG CGC CGG
 his leu pro pro arg arg arg cys arg arg phe pro phe pro arg asn val arg arg arg
 61/21 91/31
 GCG TGA TGC AAG GCT GCC TTG AGA GCA CCA GCG GCT TGA TCA TGG GCA TCG ACA GCA AGA
 ala OPA cys lys ala ala leu arg ala pro ala ala OPA ser trp ala ser thr ala arg
 121/41 151/51
 CCG CAC TGG TCG CCG AGC GCA TCA CCG GTG CCG TCG AGG AGA TC
 pro his trp ser pro ser ala ser pro val pro ser arg arg

SEQ ID N° 39B

FIGURE 39B

1/1 31/11
 GAC ACC TCC CCC CCC GCC GCC GCT GCC GCC GGT TCC CTT TCC CAA GGA ATG TCC GGC GCC
 asp thr ser pro pro ala ala ala ala ala gly ser leu ser gln gly met ser gly ala
 61/21 91/31
 GGG CGT GAT GCA AGG CTG CCT TGA GAG CAC CAG CGG CTT GAT CAT GGG CAT CGA CAG CAA
 gly arg asp ala arg leu pro OPA glu his gln arg leu asp his gly his arg gln gln
 121/41 151/51
 GAC CGC ACT GGT CGC CGA GCG CAT CAC CGG TGC CGT CGA GGA GAT C
 asp arg thr gly arg arg ala his his arg cys arg arg gly asp

SEQ ID N° 39C

FIGURE 39C

125/185

Séquence codante Rv3006 prédite par Cole et al., 1998 (Nature 393 537-544) et contenant seq39A

```

1/1                               31/11
ATG TGG ACA ACG CGG TTG GTT CGA TCC GGA CTC GCC GCG CTG TGC GCG GCA GTG CTG GTA
Met trp thr thr arg leu val arg ser gly leu ala ala leu cys ala ala val leu val
61/21                               91/31
TCG AGC GGC TGC GCA CGG TTC AAC GAC GCT CAA TCT CAG CCG TTC ACC ACC GAA CCG GAG
ser ser gly cys ala arg phe asn asp ala gln ser gln pro phe thr thr glu pro glu
121/41                               151/51
CTG CGG CCC CAA CCC AGC TCG ACA CCT CCC CCC CCG CCG CCG CTG CCG CCG GTT CCC TTT
leu arg pro gln pro ser ser thr pro pro pro pro pro pro leu pro pro val pro phe
181/61                               211/71
CCC AAG GAA TGT CCG GCG CCG GGC GTG ATG CAA GGC TGC CTT GAG AGC ACC AGC GGC TTG
pro lys glu cys pro ala pro gly val met gln gly cys leu glu ser thr ser gly leu
241/81                               271/91
ATC ATG GGC ATC GAC AGC AAG ACC GCA CTG GTC GCC GAG CGC ATC ACC GGT GCC GTC GAG
ile met gly ile asp ser lys thr ala leu val ala glu arg ile thr gly ala val glu
301/101                               331/111
GAG ATC TCT ATC AGC GCC GAG CCG AAG GTA AAG ACG GTC ATC CCC GTG GAT CCT GCC GGT
glu ile ser ile ser ala glu pro lys val lys thr val ile pro val asp pro ala gly
361/121                               391/131
GAC GGT GGC TTG ATG GAC ATT GTG CTG TCG CCC ACC TAC TCG CAA GAC CGG CTG ATG TAC
asp gly gly leu met asp ile val leu ser pro thr tyr ser gln asp arg leu met tyr
421/141                               451/151
GCC TAC ATC AGC ACG CCC ACC GAC AAC CGG GTG GTG CGA GTG GCC GAC GGC GAC ATC CCC
ala tyr ile ser thr pro thr asp asn arg val val arg val ala asp gly asp ile pro
481/161                               511/171
AAG GAC ATC CTG ACC GGC ATC CCC AAA GGT GCT GCC GGT AAC ACC GGG GCG CTG ATC TTC
lys asp ile leu thr gly ile pro lys gly ala ala gly asn thr gly ala leu ile phe
541/181                               571/191
ACC AGT CCC ACC ACG CTG GTC GTG ATG ACC GGG GAT GCT GGC GAC CCG GCG TTG GCC GCC
thr ser pro thr thr leu val val met thr gly asp ala gly asp pro ala leu ala ala
601/201                               631/211
GAT CCC CAA TCG TTG GCC GGT AAG GTC CTG CGT ATC GAA CAG CCC ACC ACC ATC GGC CAG
asp pro gln ser leu ala gly lys val leu arg ile glu gln pro thr thr ile gly gln
661/221                               691/231
ACG CCG CCG ACG ACG GCG CTG TCT GGC ATC GGC TCC GGC GGC GGC TTG TGC ATC GAT CCG
thr pro pro thr thr ala leu ser gly ile gly ser gly gly gly leu cys ile asp pro
721/241                               751/251
GTC GAC GGC TCG CTA TAT GTC GCC GAC CGC ACG CCA ACG GCG GAC CGA TTG CAG CGC ATC
val asp gly ser leu tyr val ala asp arg thr pro thr ala asp arg leu gln arg ile
781/261                               811/271
ACC AAG AAC TCG GAG GTC TCT ACG GTA TGG ACC TGG CCG GAC AAG CCC GGC GTG GCC GGG
thr lys asn ser glu val ser thr val trp thr trp pro asp lys pro gly val ala gly
841/281                               871/291
TGT GCC GCG ATG GAC GGC ACC GTG CTG GTC AAC CTG ATT AAT ACC AAA CTG ACG GTG GCG
cys ala ala met asp gly thr val leu val asn leu ile asn thr lys leu thr val ala
901/301                               931/311
GTC CGG CTC GCG CCG TCG ACC GGT GCG GTC ACC GGA GAA CCC GAC GTT GTC CGC AAA GAC
val arg leu ala pro ser thr gly ala val thr gly glu pro asp val val arg lys asp
961/321                               991/331
ACT CAT GCG CAT GCG TGG GCA TTA CGG ATG TCG CCG GAC GGC AAC GTC TGG GGA GCC ACC
thr his ala his ala trp ala leu arg met ser pro asp gly asn val trp gly ala thr
1021/341                               1051/351
GTC AAC AAG ACC GCC GGC GAC GCC GAG AAG CTC GAC GAT GTG GTG TTC CCG CTG TTC CCG
val asn lys thr ala gly asp ala glu lys leu asp asp val val phe pro leu phe pro
1081/361                               1111/371
CAG GGT GGC GGC TTC CCG CGC AAC AAC GAC GAC AAG ACC TGA
gln gly gly gly phe pro arg asn asn asp asp lys thr OPA

```

SEQ ID N° 39D

FIGURE 39D
FEUILLE DE REMPLACEMENT (REGLE 26)

126/185

ORF d'après Cole et al., 1998 (Nature 393 537-544) et contenant Rv3006

```

1/1                                31/11
TAA GGC CAT TTA GTG CCG AAT TGG GGA TTT GAG CGG CGC TTT CGC CAG ACA ATC CGC ACA
OCH gly his leu val pro asn trp gly phe glu arg arg phe arg gln thr ile arg thr
61/21                                91/31
TTG ACC CTG ACC AGC CCA CCA AAA GGC CCC AAT TGG GCC GCC ATG CCG ACA GTG CGC ACC
leu thr leu thr ser pro pro lys gly pro asn trp ala ala met pro thr val arg thr
121/41                                151/51
CCG GCA GGT GGC GGC GAT GCC CAC AAT GTC CGT AGC CTG TCG GTC ATG TGG ACA ACG CGG
pro ala gly gly gly asp ala his asn val arg ser leu ser val met trp thr thr arg
181/61                                211/71
TTG GTT CGA TCC GGA CTC GCC GCG CTG TGC GCG GCA GTG CTG GTA TCG AGC GGC TGC GCA
leu val arg ser gly leu ala ala leu cys ala ala val leu val ser ser gly cys ala
241/81                                271/91
CGG TTC AAC GAC GCT CAA TCT CAG CCG TTC ACC ACC GAA CCG GAG CTG CCG CCC CAA CCC
arg phe asn asp ala gln ser gln pro phe thr thr glu pro glu leu arg pro gln pro
301/101                                331/111
AGC TCG ACA CCT CCC CCC CCG CCG CCG CTG CCG CCG GTT CCC TTT CCC AAG GAA TGT CCG
ser ser thr pro pro pro pro pro pro leu pro pro val pro phe pro lys glu cys pro
361/121                                391/131
GCG CCG GGC GTG ATG CAA GGC TGC CTT GAG AGC ACC AGC GGC TTG ATC ATG GGC ATC GAC
ala pro gly val met gln gly cys leu glu ser thr ser gly leu ile met gly ile asp
421/141                                451/151
AGC AAG ACC GCA CTG GTC GCC GAG CGC ATC ACC GGT GCC GTC GAG GAG ATC TCT ATC AGC
ser lys thr ala leu val ala glu arg ile thr gly ala val glu glu ile ser ile ser
481/161                                511/171
GCC GAG CCG AAG GTA AAG ACG GTC ATC CCC GTG GAT CCT GCC GGT GAC GGT GGC TTG ATG
ala glu pro lys val lys thr val ile pro val asp pro ala gly asp gly gly leu met
541/181                                571/191
GAC ATT GTG CTG TCG CCC ACC TAC TCG CAA GAC CGG CTG ATG TAC GCC TAC ATC AGC ACG
asp ile val leu ser pro thr tyr ser gln asp arg leu met tyr ala tyr ile ser thr
601/201                                631/211
CCC ACC GAC AAC CGG GTG GTG CGA GTG GCC GAC GGC GAC ATC CCC AAG GAC ATC CTG ACC
pro thr asp asn arg val val arg val ala asp gly asp ile pro lys asp ile leu thr
661/221                                691/231
GGC ATC CCC AAA GGT GCT GCC GGT AAC ACC GGG GCG CTG ATC TTC ACC AGT CCC ACC ACG
gly ile pro lys gly ala ala gly asn thr gly ala leu ile phe thr ser pro thr thr
721/241                                751/251
CTG GTC GTG ATG ACC GGG GAT GCT GGC GAC CCG GCG TTG GCC GCC GAT CCC CAA TCG TTG
leu val val met thr gly asp ala gly asp pro ala leu ala ala asp pro gln ser leu
781/261                                811/271
GCC GGT AAG GTC CTG CGT ATC GAA CAG CCC ACC ACC ATC GGC CAG ACG CCG CCG ACG ACG
ala gly lys val leu arg ile glu gln pro thr thr ile gly gln thr pro pro thr thr
841/281                                871/291
GCG CTG TCT GGC ATC GGC TCC GGC GGC GGC TTG TGC ATC GAT CCG GTC GAC GGC TCG CTA
ala leu ser gly ile gly ser gly gly gly leu cys ile asp pro val asp gly ser leu
901/301                                931/311
TAT GTC GCC GAC CGC ACG CCA ACG GCG GAC CGA TTG CAG CGC ATC ACC AAG AAC TCG GAG
tyr val ala asp arg thr pro thr ala asp arg leu gln arg ile thr lys asn ser glu

```

SEQ ID N° 39F

FIGURE 39F

FEUILLE DE REMPLACEMENT (REGLE 26)

127/185

961/321
 GTC TCT ACG GTA TGG ACC TGG CCG GAC AAG CCC GGC GTG GCC GGG TGT GCC GCG ATG GAC
 val ser thr val trp thr trp pro asp lys pro gly val ala gly cys ala ala met asp
 1021/341
 GGC ACC GTG CTG GTC AAC CTG ATT AAT ACC AAA CTG ACG GTG GCG GTC CGG CTC GCG CCG
 gly thr val leu val asn leu ile asn thr lys leu thr val ala val arg leu ala pro
 1081/361
 TCG ACC GGT GCG GTC ACC GGA GAA CCC GAC GTT GTC CGC AAA GAC ACT CAT GCG CAT GCG
 ser thr gly ala val thr gly glu pro asp val val arg lys asp thr his ala his ala
 1141/381
 TGG GCA TTA CGG ATG TCG CCG GAC GGC AAC GTC TGG GGA GCC ACC GTC AAC AAG ACC GCC
 trp ala leu arg met ser pro asp gly asn val trp gly ala thr val asn lys thr ala
 1201/401
 GGC GAC GCC GAG AAG CTC GAC GAT GTG GTG TTC CCG CTG TTC CCG CAG GGT GGC GGC TTC
 gly asp ala glu lys leu asp asp val val phe pro leu phe pro gln gly gly gly phe
 1261/421
 CCG CGC AAC AAC GAC GAC AAG ACC TGA
 pro arg asn asn asp asp lys thr OPA

SEQ ID N° 39F (suite)

FIGURE 39F (suite)

1/1
 GAA GGC CTT GTT GAG CCG GCG CAC GAA AAC GAT CGT TGT GTG TAC ATT GGT GTG TAT GGC
 glu gly leu val glu pro ala his glu asn asp arg cys val tyr ile gly val tyr gly
 61/21
 TCG GTT GAA CGT GTA TGT GCC CGA CGA ATT GGC GGA GCG CGC CAG GGC GCG GGG CTT GAA
 ser val glu arg val cys ala arg arg ile gly gly ala arg gln gly ala gly leu glu
 121/41
 CGT CTC GGC GCT GAC TCA GGC CGC GAT CAG TGC CGA GTT GGA GAA CTC CGC AAC CGA TGC
 arg leu gly ala asp ser gly arg asp gln cys arg val gly glu leu arg asn arg cys
 181/61
 GTG GCT TGA GGG GTT GGA ACC CAG AAG CAC CGG CGC TCG GCA TGA TGA CGT GCT GGG TGC
 val ala OPA gly val gly thr gln lys his arg arg ser ala OPA OPA arg ala gly cys
 241/81
 GAT CGA TGC CGC TCG CGA TGA GTT CGA AGC GTG AGA GCA TCG CCC ACT TCG CCG CCG GAG
 asp arg cys arg ser arg OPA val arg ser val arg ala ser pro thr ser pro pro glu
 301/101
 CAG GTG GTC GTC GAC GCG AGT GCC ATG GTG GAT C
 gln val val val asp ala ser ala met val asp

SEQ ID N° 40A

FIGURE 40A

128/185

```

1/1                                31/11
AAG GCC TTG TTG AGC CGG CGC ACG AAA ACG ATC GTT GTG TGT ACA TTG GTG TGT ATG GCT
lys ala leu leu ser arg arg thr lys thr ile val val cys thr leu val cys met ala
61/21                                91/31
CGG TTG AAC GTG TAT GTG CCC GAC GAA TTG GCG GAG CGC GCC AGG GCG CGG GGC TTG AAC
arg leu asn val tyr val pro asp glu leu ala glu arg ala arg ala arg gly leu asn
121/41                                151/51
GTC TCG GCG CTG ACT CAG GCC GCG ATC AGT GCC GAG TTG GAG AAC TCC GCA ACC GAT GCG
val ser ala leu thr gln ala ala ile ser ala glu leu glu asn ser ala thr asp ala
181/61                                211/71
TGG CTT GAG GGG TTG GAA CCC AGA AGC ACC GGC GCT CGG CAT GAT GAC GTG CTG GGT GCG
trp leu glu gly leu glu pro arg ser thr gly ala arg his asp asp val leu gly ala
241/81                                271/91
ATC GAT GCC GCT CGC GAT GAG TTC GAA GCG TGA GAG CAT CGC CCA CTT CGC CGC CGG AGC
ile asp ala ala arg asp glu phe glu ala OPA glu his arg pro leu arg arg arg ser
301/101                                331/111
AGG TGG TCG TCG ACG CGA GTG CCA TGG TGG ATC
arg trp ser ser thr arg val pro trp trp ile

```

SEQ ID N° 40B

FIGURE 40B

```

1/1                                31/11
AGG CCT TGT TGA GCC GGC GCA CGA AAA CGA TCG TTG TGT GTA CAT TGG TGT GTA TGG CTC
arg pro cys OPA ala gly ala arg lys arg ser leu cys val his trp cys val trp leu
61/21                                91/31
GGT TGA ACG TGT ATG TGC CCG ACG AAT TGG CGG AGC GCG CCA GGG CGC GGG GCT TGA ACG
gly OPA thr cys met cys pro thr asn trp arg ser ala pro gly arg gly ala OPA thr
121/41                                151/51
TCT CGG CGC TGA CTC AGG CCG CGA TCA GTG CCG AGT TGG AGA ACT CCG CAA CCG ATG CGT
ser arg arg OPA leu arg pro arg ser val pro ser trp arg thr pro gln pro met arg
181/61                                211/71
GGC TTG AGG GGT TGG AAC CCA GAA GCA CCG GCG CTC GGC ATG ATG ACG TGC TGG GTG CGA
gly leu arg gly trp asn pro glu ala pro ala leu gly met met thr cys trp val arg
241/81                                271/91
TCG ATG CCG CTC GCG ATG AGT TCG AAG CGT GAG AGC ATC GCC CAC TTC GCC GCC GGA GCA
ser met pro leu ala met ser ser lys arg glu ser ile ala his phe ala ala gly ala
301/101
GGT GGT CGT CGA CGC GAG TGC CAT GGT GGA TC
gly gly arg arg arg glu cys his gly gly

```

SEQ ID N° 40C

FIGURE 40C

129/185

Séquence codante Rv0549c prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq40A

```

1/1                               31/11
gtg aga gca tcg ccc act tcg ccg ccg gag cag gtg gtc gtc gac gcg agt gcc atg gtg
val arg ala ser pro thr ser pro pro glu gln val val val asp ala ser ala met val
61/21                               91/31
gat cta ctg gct cgc act agc gat ccg tgc tct gcg gtg cgc gcg ccg ctg gct ccg acc
asp leu leu ala arg thr ser asp arg cys ser ala val arg ala arg leu ala arg thr
121/41                             151/51
gcg atg cac gcg ccg gcg cac ttc gat gca gag gtg ttg tcg gcg ctg ggg cgc atg cag
ala met his ala pro ala his phe asp ala glu val leu ser ala leu gly arg met gln
181/61                             211/71
cgc gcc gcc gca ctc acc gtt gcc tat gtc gat gcg gca ctg gag gag ttg cga cag gtg
arg ala gly ala leu thr val ala tyr val asp ala ala leu glu glu leu arg gln val
241/81                             271/91
ccg gtg act cga cac ggt ctt tcg tcg ctg ctt gct gga gcg tgg tcg cgc cgc gac acc
pro val thr arg his gly leu ser ser leu leu ala gly ala trp ser arg arg asp thr
301/101                             331/111
ctc cgc ctg acc gat gcc ctc tac gtc gag ctg gcc gaa acg gca ggt ctg gtg ttg ttg
leu arg leu thr asp ala leu tyr val glu leu ala glu thr ala gly leu val leu leu
361/121                             391/131
acc acc gac gaa aga ttg gca cgc gcc tgg ccc tcg gct cac gcc atc ggc tga
thr thr asp glu arg leu ala arg ala trp pro ser ala his ala ile gly OPA

```

SEQ ID N° 40D

FIGURE 40D

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0549c

```

1/1                               31/11
tga gtt cga agc gtg aga gca tcg ccc act tcg ccg ccg gag cag gtg gtc gtc gac gcg
OPA val arg ser val arg ala ser pro thr ser pro pro glu gln val val val asp ala
61/21                               91/31
agt gcc atg gtg gat cta ctg gct cgc act agc gat ccg tgc tct gcg gtg cgc gcg ccg
ser ala met val asp leu leu ala arg thr ser asp arg cys ser ala val arg ala arg
121/41                             151/51
ctg gct ccg acc gcg atg cac gcg ccg gcg cac ttc gat gca gag gtg ttg tcg gcg ctg
leu ala arg thr ala met his ala pro ala his phe asp ala glu val leu ser ala leu
181/61                             211/71
ggg cgc atg cag cgc gcc gcc gca ctc acc gtt gcc tat gtc gat gcg gca ctg gag gag
gly arg met gln arg ala gly ala leu thr val ala tyr val asp ala ala leu glu glu
241/81                             271/91
ttg cga cag gtg ccg gtg act cga cac ggt ctt tcg tcg ctg ctt gct gga gcg tgg tcg
leu arg gln val pro val thr arg his gly leu ser ser leu leu ala gly ala trp ser
301/101                             331/111
cgc cgc gac acc ctc cgc ctg acc gat gcc ctc tac gtc gag ctg gcc gaa acg gca ggt
arg arg asp thr leu arg leu thr asp ala leu tyr val glu leu ala glu thr ala gly
361/121                             391/131
ctg gtg ttg ttg acc acc gac gaa aga ttg gca cgc gcc tgg ccc tcg gct cac gcc atc
leu val leu leu thr thr asp glu arg leu ala arg ala trp pro ser ala his ala ile
421/141
ggc tga
gly OPA

```

SEQ ID N° 40F

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 40F

130/185

```

1/1                               31/11
CCT GGC CGG GAC GCC TAC GTG TAG CCC GCG GCT AGC ACA GGA TAG CCA TTG TTG TGC GGT
pro gly arg asp ala tyr val AMB pro ala ala ser thr gly AMB pro leu leu cys gly
61/21                               91/31
AGC GCC AAA ACG ATC AGC CCT TCG CGG ACA TGT CAG CAC CCG CCT TGG CCG GGA GAG CGG
ser ala lys thr ile ser pro ser arg thr cys gln his pro pro trp pro gly glu arg
121/41                               151/51
CGT CGT GAC CGT GCT GTC ACC ACG TCT GGT TAG GCT CGG GGC GCG GGC TGG CGC GGA GGA
arg arg asp arg ala val thr thr ser gly AMB ala arg gly ala gly trp arg gly gly
181/61                               211/71
GGT GTG TTG CGG AGG AGG TGT GTT GTA GTG GGG ACG GCG GAT CGG CCG TTG GAC GCC TCG
gly val leu arg arg arg cys val val val gly thr ala asp arg pro leu asp ala ser
241/81                               271/91
GCC TTG CGG GAC TGG GCA CAC GCC GTC GTC AGC GAT C
ala leu arg asp trp ala his ala val val ser asp

```

SEQ ID N° 41A

FIGURE 41A

```

1/1                               31/11
CTG GCC GGG ACG CCT ACG TGT AGC CCG CGG CTA GCA CAG GAT AGC CAT TGT TGT GCG GTA
leu ala gly thr pro thr cys ser pro arg leu ala gln asp ser his cys cys ala val
61/21                               91/31
GCG CCA AAA CGA TCA GCC CTT CGC GGA CAT GTC AGC ACC CGC CTT GGC CGG GAG AGC GGC
ala pro lys arg ser ala leu arg gly his val ser thr arg leu gly arg glu ser gly
121/41                               151/51
GTC GTG ACC GTG CTG TCA CCA CGT CTG GTT AGG CTC GGG GCG CGG GCT GGC GCG GAG GAG
val val thr val leu ser pro arg leu val arg leu gly ala arg ala gly ala glu glu
181/61                               211/71
GTG TGT TGC GGA GGA GGT GTG TTG TAG TGG GGA CGG CGG ATC GGC CGT TGG ACG CCT CGG
val cys cys gly gly gly val leu AMB trp gly arg arg ile gly arg trp thr pro arg
241/81                               271/91
CCT TGC GGG ACT GGG CAC ACG CCG TCG TCA GCG ATC
pro cys gly thr gly his thr pro ser ser ala ile

```

SEQ ID N° 41B

FIGURE 41B

131/185

```

1/1                               31/11
TGG CCG GGA CGC CTA CGT GTA GCC CGC GGC TAG CAC AGG ATA GCC ATT GTT GTG CGG TAG
trp pro gly arg leu arg val ala arg gly AMB his arg ile ala ile val val arg AMB
61/21                               91/31
CGC CAA AAC GAT CAG CCC TTC GCG GAC ATG TCA GCA CCC GCC TTG GCC GGG AGA GCG GCG
arg gln asn asp gln pro phe ala asp met ser ala pro ala leu ala gly arg ala ala
121/41                               151/51
TCG TGA CCG TGC TGT CAC CAC GTC TGG TTA GGC TCG GGG CGC GGG CTG GCG CGG AGG AGG
ser OPA pro cys cys his his val trp leu gly ser gly arg gly leu ala arg arg arg
181/61                               211/71
TGT GTT GCG GAG GAG GTG TGT TGT AGT GGG GAC GGC GGA TCG GCC GTT GGA CGC CTC GGC
cys val ala glu glu val cys cys ser gly asp gly gly ser ala val gly arg leu gly
241/81                               271/91
CTT GCG GGA CTG GGC ACA CGC CGT CGT CAG CGA TC
leu ala gly leu gly thr arg arg arg gln arg

```

SEQ ID N° 41C

FIGURE 41C

Séquence codante Rv2975c prédite par Cole et al, 1998 (Nature 393: 537-544) et contenant seq41A

```

1/1                               31/11
gtg ggg acg gcg gat cgg ccg ttg gac gcc tcg gcc ttg cgg gac tgg gca cac gcc gtc
val gly thr ala asp arg pro leu asp ala ser ala leu arg asp trp ala his ala val
61/21                               91/31
gtc agc gat ctg atc ctc cac atc gac gag atc aac cgg ctc aat gtg ttc ccg gtc gct
val ser asp leu ile leu his ile asp glu ile asn arg leu asn val phe pro val ala
121/41                               151/51
gac tcc gat acc ggc gtc aac atg ctg ttc acc atg cgt gcc gcg gtc gta gaa gct gat
asp ser asp thr gly val asn met leu phe thr met arg ala ala val val glu ala asp
181/61                               211/71
ttg cac gcg aat tcg cag gct gac gcc gaa gac gtg gcg cgg gtt gcg gcc gct ctc gcg
leu his ala asn ser gln ala asp ala glu asp val ala arg val ala ala ala leu ala
241/81
gcc ggc gcg cgt tga
ala gly ala arg OPA

```

SEQ ID N° 41D

FIGURE 41D

132/185

ORF d'après Cole et al, 1998 (Nature 393: 537-544) et contenant Rv2975c

```

1/1                               31/11
tag gct cgg ggc gcg ggc tgg cgc gga gga ggt gtg ttg cgg agg agg tgt gtt gta gtg
AMB ala arg gly ala gly trp arg gly gly gly val leu arg arg arg cys val val val
61/21                               91/31
ggg acg gcg gat cgg ccg ttg gac gcc tcg gcc ttg cgg gac tgg gca cac gcc gtc gtc
gly thr ala asp arg pro leu asp ala ser ala leu arg asp trp ala his ala val val
121/41                               151/51
agc gat ctg atc ctc cac atc gac gag atc aac cgg ctc aat gtg ttc ccg gtc gct gac
ser asp leu ile leu his ile asp glu ile asn arg leu asn val phe pro val ala asp
181/61                               211/71
tcc gat acc ggc gtc aac atg ctg ttc acc atg cgt gcc gcg gtc gta gaa gct gat ttg
ser asp thr gly val asn met leu phe thr met arg ala ala val val glu ala asp leu
241/81                               271/91
cac gcg aat tcg cag gct gac gcc gaa gac gtg gcg cgg gtt gcg gcc gct ctc gcg gcc
his ala asn ser gln ala asp ala glu asp val ala arg val ala ala ala leu ala ala
301/101
ggc gcg cgt tga
gly ala arg OPA

```

SEQ ID N° 41F

FIGURE 41F

séquence Rv 2974C prédite par Cole et al. (Nature 393:537-544) et pouvant être dans la même phase de lecture que Seq41D. Le séquençage de cette région fait apparaître dans un cas sur trois une délétion de deux nucléotides mettant en phase bservé dans

```

1/1                               31/11
ttg aac gga gct cgc ggc aac tcc ggc gtg atc ctg tcc cag atc ctg cgc ggg atc gca
leu asn gly ala arg gly asn ser gly val ile leu ser gln ile leu arg gly ile ala
61/21                               91/31
gag gtg acc gcg act gcg gcc gcc gcc tct ggc gcg gta ttg cgg gcg gtc gac gcc aac
glu val thr ala thr ala ala ala ala ser gly ala val leu arg ala val asp ala asn
121/41                               151/51
gcc ctc ggg gcc gcg ttg tgg cgc ggc gtc gag ttg gtc gtc gcg tcg atg ggt gcc gtg
ala leu gly ala ala leu trp arg gly val glu leu val val ala ser met gly gly val
181/61                               211/71
gag gtg ccg gga act atc gtc tcg gtg ctg cgg gcc gcc gcc gga gcc gtc gac cag tgc
glu val pro gly thr ile val ser val leu arg ala ala ala gly ala val asp gln cys
241/81                               271/91
gcg cac gag ggg ttg gcc ggt gcg gtc acc gcc gcc ggt gac gcg gcg gtc atc gcg ctg
ala his glu gly leu ala gly ala val thr ala ala gly asp ala ala val ile ala leu
301/101                               331/111
gaa aag acc ccc gaa cag ctt gac gtg ctc gcc gat gcg ggc gcg gtg gac gcc gcc gga
glu lys thr pro glu gln leu asp val leu ala asp ala gly ala val asp ala gly gly

```

SEQ ID N° 41S

FIGURE 41S

FEUILLE DE REMPLACEMENT (REGLE 26)

133/185

361/121
 cgg ggc ctg ctg gtt ctg ctg gac gcg ttg cgc tcc acc atc tgc ggg cag gca cct gcc
 arg gly leu leu val leu leu asp ala leu arg ser thr ile cys gly gln ala pro ala
 421/141
 cgg ggc gtc tac gaa ccc tcg ccg cgc gcg ttg ccg acc gac acg gct acc caa cgc ccc
 arg ala val tyr glu pro ser pro arg ala leu pro thr asp thr ala thr gln arg pro
 481/161
 gcc ccg caa ttc gag gtg atg tat ctg ttg gcg gta tgt gat gct gca gcg gcg gac cag
 ala pro gln phe glu val met tyr leu leu ala val cys asp ala ala ala asp gln
 541/181
 ttg cgg gat cga ctc aag gaa ttg ggt gag tcg gtg gcc atc gcc gct gct ccg ccc gac
 leu arg asp arg leu lys glu leu gly glu ser val ala ile ala ala ala pro pro asp
 601/201
 agc tac tcc gta cac gtc cac acc gac gac gcc ggt gcc gcc gtg gaa gcc gga ttg gcg
 ser tyr ser val his val his thr asp asp ala gly ala val glu ala gly leu ala
 661/221
 gtg ggg cga gtt agc cgg atc gtg atc tcg gcg ctc ggt tcc ggg acc agc gga ttg ccg
 val gly arg val ser arg ile val ile ser ala leu gly ser gly thr ser gly leu pro
 721/241
 gcc ggt ggc tgg acg cgg ggc cgc gcc gtg ctg gcg gtc gtc gac ggc gac ggt gcc gcc
 ala gly gly trp thr arg gly arg ala val leu ala val val asp gly asp gly ala ala
 781/261
 gag ctg ttc gcc ggg gag ggc gcc tgc gtg ctg cga ccg ggt cca gac gcc gtg aca ccg
 glu leu phe ala gly glu gly ala cys val leu arg pro gly pro asp ala val thr pro
 841/281
 gcc gcc gat atc agt gcc cac cag ctg gtg cgg gcc gtg gta gac acc ggc gcc gcg cac
 ala ala asp ile ser ala his gln leu val arg ala val val asp thr gly ala ala his
 901/301
 gtg atg gtg ctg ccc aat ggc tat gtg gcc gcc gaa gaa ctg gtg gcc ggg tgt acc gcg
 val met val leu pro asn gly tyr val ala ala glu glu leu val ala gly cys thr ala
 961/321
 gcg atc ggc tgg ggc gtc gac gtg gta ccc gtg ccg acc gga tcg atg gtg cag ggg ttg
 ala ile gly trp gly val asp val val pro val pro thr gly ser met val gln gly leu
 1021/341
 gcc gcg ctg gcc gtg cat gac gcg gcc cgc cag gcc gtc gac gac ggc tac agc atg gcc
 ala ala leu ala val his asp ala ala arg gln ala val asp asp gly tyr ser met ala
 1081/361
 cgt gcc gcc ggt gct tcc cgg cac gga tcg gtg cgc att gcc acc caa aag gcg ctg acc
 arg ala ala gly ala ser arg his gly ser val arg ile ala thr gln lys ala leu thr
 1141/381
 tgg gcc ggt acc tgc aag ccg ggc gac ggt ctg ggt atc gcg ggc gac gag gtg ctg atc
 trp ala gly thr cys lys pro gly asp gly leu gly ile ala gly asp glu val leu ile
 1201/401
 gtc gcc gac gat gtc gcc gcg gcg gcc atc ggt ctg gtc gac ctg ttg ttg gca tcg gga
 val ala asp asp val ala ala ala ala ile gly leu val asp leu leu leu ala ser gly
 1261/421
 ggc gat ctg gtg acg gtg cta att ggc gcc ggc gta acc gaa gac gtg gct gtc gtc ctg
 gly asp leu val thr val leu ile gly ala gly val thr glu asp val ala val val leu
 1321/441
 gaa cgg cat gtg cac gac cac cat cca ggc acc gag ctg gtc tcc tac cgc acc gga cac
 glu arg his val his asp his his pro gly thr glu leu val ser tyr arg thr gly his
 1381/461
 cgc ggc gac gcg ctg ctg atc ggg gtc gag tag
 arg gly asp ala leu leu ile gly val glu AMB

SEQ ID N° 41S (suite)

FIGURE 41S (suite)
 FEUILLE DE REMPLACEMENT (REGLE 26)

134/185

Seq41T comprenant seq 41F et seq 41S

1/1 31/11
tta ggc tgc ggg cgc ggg ctg gcg cgg agg agg tgt gtt gcg gag gag gtg tgt tgt agt
leu gly ser gly arg gly leu ala arg arg arg cys val ala glu glu val cys cys ser
AMB ala arg gly ala gly trp arg gly gly gly val leu arg arg arg cys val val val
arg leu gly ala arg ala gly ala glu glu val cys cys gly gly gly val leu AMB trp
61/21 91/31
ggg gac ggc gga tgc gcc gtt gga cgc ctc ggc ctt gcg gga ctg ggc aca cgc cgt cgt
gly asp gly gly ser ala val gly arg leu gly leu ala gly leu gly thr arg arg arg
gly thr ala asp arg pro leu asp ala ser ala leu arg asp trp ala his ala val val
gly arg arg ile gly arg trp thr pro arg pro cys gly thr gly his thr pro ser ser
121/41 151/51
cag cga tct gat cct cca cat cga cga gat caa ccg gct caa tgt gtt ccc ggt cgc tga
gln arg ser asp pro pro his arg arg asp gln pro ala gln cys val pro gly arg OPA
ser asp leu ile leu his ile asp glu ile asn arg leu asn val phe pro val ala asp
ala ile OPA ser ser thr ser thr arg ser thr gly ser met cys ser arg ser leu thr
181/61 211/71
ctc cga tac cgg cgt caa cat gct gtt cac cat gcg tgc cgc ggt cgt aga agc tga ttt
leu arg tyr arg arg gln his ala val his his ala cys arg gly arg arg ser OPA phe
ser asp thr gly val asn met leu phe thr met arg ala ala val val glu ala asp leu
pro ile pro ala ser thr cys cys ser pro cys val pro arg ser AMB lys leu ile cys
241/81 271/91
gca cgc gaa ttc gca ggc tga cgc cga aga cgt ggc gcg ggt tgc ggc cgc tct cgc ggc
ala arg glu phe ala gly OPA arg arg arg arg gly ala gly cys gly arg ser arg gly
his ala asn ser gln ala asp ala glu asp val ala arg val ala ala leu ala ala
thr arg ile arg arg leu thr pro lys thr trp arg gly leu arg pro leu ser arg pro
301/101 331/111
cgg cgc gcg ttg aac gga gct cgc ggc aac tcc ggc gtg atc ctg tcc cag atc ctg cgc
arg arg ala leu asn gly ala arg gly asn ser gly val ile leu ser gln ile leu arg
gly ala arg OPA thr glu leu ala ala thr pro ala OPA ser cys pro arg ser cys ala
ala arg val glu arg ser ser arg gln leu arg arg asp pro val pro asp pro ala arg
361/121 391/131
ggg atc gca gag gtg acc gcg act gcg gcc gcc gcc tct ggc gcg gta ttg cgg gcg gtc
gly ile ala glu val thr ala thr ala ala ala ala ser gly ala val leu arg ala val
gly ser gln arg OPA pro arg leu arg pro pro pro leu ala arg tyr cys gly arg ser
asp arg arg gly asp arg asp cys gly arg arg leu trp arg gly ile ala gly gly arg
421/141 451/151
gac gcc aac gcc ctc ggg gcc gcg ttg tgg cgc ggc gtc gag ttg gtc gtc gcg tgc atg
asp ala asn ala leu gly ala ala leu trp arg gly val glu leu val val ala ser met
thr pro thr pro ser gly pro arg cys gly ala ala ser ser trp ser ser arg arg trp
arg gln arg pro arg gly arg val val ala arg arg val gly arg arg val asp gly
481/161 511/171
ggt ggc gtg gag gtg ccg gga act atc gtc tgc gtg ctg cgg gcc gcc gcc gga gcc gtc
gly gly val glu val pro gly thr ile val ser val leu arg ala ala ala gly ala val
val ala trp arg cys arg glu leu ser ser arg cys cys gly pro pro pro glu pro ser
trp arg gly gly ala gly asn tyr arg leu gly ala ala gly arg arg ser arg arg
541/181 571/191
gac cag tgc gcg cac gag ggg ttg gcc ggt gcg gtc acc gcc gcc ggt gac gcg gcg gtc
asp gln cys ala his glu gly leu ala gly ala val thr ala ala gly asp ala ala val
thr ser ala arg thr arg gly trp pro val arg ser pro pro pro val thr arg arg ser
pro val arg ala arg gly val gly arg cys gly his arg arg arg OPA arg gly gly his
601/201 631/211
atc gcg ctg gaa aag acc ccc gaa cag ctt gac gtg ctc gcc gat gcg ggc gcg gtg gac
ile ala leu glu lys thr pro glu gln leu asp val leu ala asp ala gly ala val asp
ser arg trp lys arg pro pro asn ser leu thr cys ser pro met arg ala arg trp thr
arg ala gly lys asp pro arg thr ala OPA arg ala arg arg cys gly arg gly arg
661/221 691/231
gcc ggc gga cgg ggc ctg ctg gtt ctg ctg gac gcg ttg cgc tcc acc atc tgc ggg cag
ala gly gly arg gly leu leu val leu leu asp ala leu arg ser thr ile cys gly gln
pro ala asp gly ala cys trp phe cys trp thr arg cys ala pro pro ser ala gly arg
arg arg thr gly pro ala gly ser ala gly arg val ala leu his his leu arg ala gly

SEQ ID N° 41T

FEUILLE DE REMPLACEMENT (REGLE 26)

135/185

721/241
gca cct gcc cgg gcg gtc tac gaa ccc tcg ccg cgc gcg ttg ccg acc gac acg gct acc
ala pro ala arg ala val tyr glu pro ser pro arg ala leu pro thr asp thr ala thr
his leu pro gly arg ser thr asn pro arg arg ala arg cys arg pro thr arg leu pro
thr cys pro gly gly leu arg thr leu ala ala arg val ala asp arg his gly tyr pro
781/261
caa cgc ccc gcc ccg caa ttc gag gtg atg tat ctg ttg gcg gta tgt gat gct gca gcg
gln arg pro ala pro gln phe glu val met tyr leu leu ala val cys asp ala ala ala
asn ala pro pro arg asn ser arg OPA cys ile cys trp arg tyr val met leu gln arg
thr pro arg pro ala ile arg gly asp val ser val gly gly met OPA cys cys ser gly
841/281
gcg gac cag ttg ccg gat cga ctc aag gaa ttg ggt gag tcg gtg gcc atc gcc gct gct
ala asp gln leu arg asp arg leu lys glu leu gly glu ser val ala ile ala ala ala
arg thr ser cys gly ile asp ser arg asn trp val ser arg trp pro ser pro leu leu
gly pro val ala gly ser thr gln gly ile gly OPA val gly gly his arg arg cys ser
901/301
ccg ccc gac agc tac tcc gta cac gtc cac acc gac gac gcc ggt gcc gcc gtg gaa gcc
pro pro asp ser tyr ser val his val his thr asp asp ala gly ala ala val glu ala
arg pro thr ala thr pro tyr thr ser thr pro thr thr pro val pro pro trp lys pro
ala arg gln leu leu arg thr arg pro his arg arg arg cys arg arg gly ser arg
961/321
gga ttg gcg gtg ggg cga gtt agc cgg atc gtg atc tcg gcg ctc ggt tcc ggg acc agc
gly leu ala val gly arg val ser arg ile val ile ser ala leu gly ser gly thr ser
asp trp arg trp gly glu leu ala gly ser OPA ser arg arg ser val pro gly pro ala
ile gly gly ala ser AMB pro asp arg asp leu gly ala arg phe arg asp gln arg
1021/341
gga ttg ccg gcc ggt ggc tgg acg cgg ggc cgc gcc gtg ctg gcg gtc gtc gac ggc gac
gly leu pro ala gly gly trp thr arg gly arg ala val leu ala val val asp gly asp
asp cys arg pro val ala gly arg gly ala ala pro cys trp arg ser ser thr ala thr
ile ala gly arg trp leu asp ala gly pro arg arg ala gly gly arg arg arg arg
1081/361
ggt gcc gcc gag ctg ttc gcc ggg gag ggc gcc tgc gtg ctg cga ccg ggt cca gac gcc
gly ala ala glu leu phe ala gly glu gly ala cys val leu arg pro gly pro asp ala
val pro pro ser cys ser pro gly arg ala pro ala cys cys asp arg val gln thr pro
cys arg arg ala val arg arg gly gly arg leu arg ala ala thr gly ser arg arg arg
1141/381
gtg aca ccg gcc gcc gat atc agt gcc cac cag ctg gtg ccg gcc gtg gta gac acc ggc
val thr pro ala ala asp ile ser ala his gln leu val arg ala val val asp thr gly
OPA his arg pro pro ile ser val pro thr ser trp cys gly pro trp AMB thr pro ala
asp thr gly arg arg tyr gln cys pro pro ala gly ala gly arg gly arg his arg arg
1201/401
gcc gcg cac gtg atg gtg ctg ccc aat ggc tat gtg gcc gcc gaa gaa ctg gtg gcc ggg
ala ala his val met val leu pro asn gly tyr val ala ala glu glu leu val ala gly
pro arg thr OPA trp cys cys pro met ala met trp pro pro lys asn trp trp pro gly
arg ala arg asp gly ala ala gln trp leu cys gly arg arg arg thr gly gly arg val
1261/421
tgt acc gcg gcg atc ggc tgg ggc gtc gac gtg gta ccc gtg ccg acc gga tcg atg gtg
cys thr ala ala ile gly trp gly val asp val val pro val pro thr gly ser met val
val pro arg arg ser ala gly ala ser thr trp tyr pro cys arg pro asp arg trp cys
tyr arg gly asp arg leu gly arg arg arg gly thr arg ala asp arg ile asp gly ala
1321/441
cag ggg ttg gcc gcg ctg gcc gtg cat gac gcg gcc cgc cag gcc gtc gac gac ggc tac
gln gly leu ala ala leu ala val his asp ala ala arg gln ala val asp asp gly tyr
arg gly trp pro arg trp pro cys met thr arg pro ala arg pro ser thr thr ala thr
gly val gly arg ala gly arg ala OPA arg gly pro pro gly arg arg arg arg leu gln
1381/461
agc atg gcc cgt gcc gcc ggt gct tcc cgg cac gga tcg gtg cgc att gcc acc caa aag
ser met ala arg ala ala gly ala ser arg his gly ser val arg ile ala thr gln lys
ala trp pro val pro pro val leu pro gly thr asp arg cys ala leu pro pro lys arg
his gly pro cys arg arg cys phe pro ala arg ile gly ala his cys his pro lys gly

SEQ ID N° 41T (suite 1)

FIGURE 41T (suite 1)
FEUILLE DE REMPLACEMENT (RÈGLE 26)

136/185

1441/481
 gcg ctg acc tgg gcc ggt acc tgc aag ccg ggc gac ggt ctg ggt atc gcg ggc gac gag
 ala leu thr trp ala gly thr cys lys pro gly asp gly leu gly ile ala gly asp glu
 arg OPA pro gly pro val pro ala ser arg ala thr val trp val ser arg ala thr arg
 ala asp leu gly arg tyr leu gln ala gly arg arg ser gly tyr arg gly arg arg gly
 1501/501
 1471/491
 gtg ctg atc gtc gcc gac gat gtc gcc gcg gcc atc ggt ctg gtc gac ctg ttg ttg
 val leu ile val ala asp asp val ala ala ala ile gly leu val asp leu leu leu
 cys OPA ser ser pro thr met ser pro arg arg pro ser val trp ser thr cys cys trp
 ala asp arg arg arg arg cys arg arg gly gly his arg ser gly arg pro val val gly
 1531/511
 1561/521
 gca tcg gga ggc gat ctg gtg acg gtg cta att ggc gcc ggc gta acc gaa gac gtg gct
 ala ser gly gly asp leu val thr val leu ile gly ala gly val thr glu asp val ala
 his arg glu ala ile trp OPA arg cys OCH leu ala pro ala OCH pro lys thr trp leu
 ile gly arg arg ser gly asp gly ala asn trp arg arg arg asn arg arg arg gly cys
 1591/531
 1621/541
 gtc gtc ctg gaa cgg cat gtg cac gac cac cat cca gcc acc gag ctg gtc tcc tac cgc
 val val leu glu arg his val his asp his his pro gly thr glu leu val ser tyr arg
 ser ser trp asn gly met cys thr thr thr ile gln ala pro ser trp ser pro thr ala
 arg pro gly thr ala cys ala arg pro pro ser arg his arg ala gly leu leu pro his
 1651/551
 1681/561
 acc gga cac cgc ggc gac gcg ctg ctg atc ggg gtc gag tag
 thr gly his arg gly asp ala leu leu ile gly val glu AMB
 pro asp thr ala ala thr arg cys OPA ser gly ser ser
 arg thr pro arg arg arg ala ala asp arg gly arg val
 1711/571

SEQ ID N° 41T (suite 2)

FIGURE 41T (suite 2)

1/1
 GCC GGT AAC GCC GCG TCC CAG TGC TAT CCG 31/11
 ala gly asn ala ala ser gln cys tyr pro TCC GCC GGA CCG CCC GAA ACA TCA GCG GCG
 61/21
 GGC GCC CCG GTC GGC CGC GGC CGG GCT CGA 91/31
 gly ala pro val gly arg gly arg ala arg pro ala pro pro gly his gln arg pro gly
 121/41
 TAT CGA GGT GGA AGC GGA CGG TGT TGG GAT GCA CGC CCA ACT TGC CGG CGA TCG CGG CGA
 tyr arg gly gly ser gly arg cys trp asp ala arg pro thr cys arg arg ser arg arg
 181/61
 TGC TCA TCG GAA CCC GCG ACG CAC ACA ATG 211/71
 cys ser ser glu pro ala thr his thr met pro ala ala pro his asp gly ala pro pro
 241/81
 GCT CTT GCA GTG ACC TGA TGA TGA CAC TCA 271/91
 ala leu ala val thr OPA OPA OPA his ser pro pro OCH gly ser ser ala ala pro glu
 301/101
 CAA TGC AGT AAG TTT ACA CAA ACG GAC TTG TAA AAA CCT GCG GAG GTG GGG TCT ATG GCC
 gln cys ser lys phe thr gln thr asp leu OCH lys pro ala glu val gly ser met ala
 361/121
 AAC AAA CGT GGC AAT GCC GGG CAG CCT CTG CCC TTG TCG GAT C 391/131
 asn lys arg gly asn ala gly gln pro leu pro leu ser asp

SEQ ID N° 42A

FIGURE 42A

FEUILLE DE REMPLACEMENT (REGLE 26)

137/185

```

1
1/1
CCG GTA ACG CCG CGT CCC AGT GCT ATC CGT CCG CCG GAC CGC CCG AAA CAT CAG CGG CGG
pro val thr pro arg pro ser ala ile arg pro pro asp arg pro lys his gln arg arg
61/21
GCG CCC CCG TCG GCC GCG GCC GGG CTC GAC CCG CTC CAC CTG GCC ATC AGC GAC CAG GTT
ala pro arg ser ala ala ala gly leu asp pro leu his leu ala ile ser asp gln val
121/41
ATC GAG GTG GAA GCG GAC GGT GTT GGG ATG CAC GCC CAA CTT GCC GGC GAT CGC GGC GAT
ile glu val glu ala asp gly val gly met his ala gln leu ala gly asp arg gly asp
181/61
GCT CAT CGG AAC CCG CGA CGC ACA CAA TGC CCG CAG CAC CGC ACG ACG GCG CCC CAC CGG
ala his arg asn pro arg arg thr gln cys pro gln his arg thr thr ala pro his arg
241/81
CTC TTG CAG TGA CCT GAT GAT GAC ACT CAC CCC CAT AAG GCT CGT CGG CTG CGC CTG AGC
leu leu gln OPA pro asp asp asp thr his pro his lys ala arg arg leu arg leu ser
301/101
AAT GCA GTA AGT TTA CAC AAA CGG ACT TGT AAA AAC CTG CGG AGG TGG GGT CTA TGG CCA
asn ala val ser leu his lys arg thr cys lys asn leu arg arg trp gly leu trp pro
361/121
ACA AAC GTG GCA ATG CCG GGC AGC CTC TGC CCT TGT CGG ATC
thr asn val ala met pro gly ser leu cys pro cys arg ile
31/11
91/31
151/51
211/71
271/91
331/111
391/131

```

SEQ ID N° 42B

FIGURE 42B

```

1/1
CGG TAA CGC CGC GTC CCA GTG CTA TCC GTC CGC CGG ACC GCC CGA AAC ATC AGC GGC GGG
arg OCH arg arg val pro val leu ser val arg arg thr ala arg asn ile ser gly gly
61/21
CGC CCC GGT CGG CCG CGG CCG GGC TCG ACC CGC TCC ACC TGG CCA TCA GCG ACC AGG TTA
arg pro gly arg pro arg pro gly ser thr arg ser thr trp pro ser ala thr arg leu
121/41
TCG AGG TGG AAG CGG ACG GTG TTG GGA TGC ACG CCC AAC TTG CCG GCG ATC GCG GCG ATG
ser arg trp lys arg thr val leu gly cys thr pro asn leu pro ala ile ala ala met
181/61
CTC ATC GGA ACC CGC GAC GCA CAC AAT GCC CGC AGC ACC GCA CGA CGG CGC CCC ACC GGC
leu ile gly thr arg asp ala his asn ala arg ser thr ala arg arg arg pro thr gly
241/81
TCT TGC AGT GAC CTG ATG ATG ACA CTC ACC CCC ATA AGG CTC GTC GGC TGC GCC TGA GCA
ser cys ser asp leu met met thr leu thr pro ile arg leu val gly cys ala OPA ala
301/101
ATG CAG TAA GTT TAC ACA AAC GGA CTT GTA AAA ACC TGC GGA GGT GGG GTC TAT GGC CAA
met gln OCH val tyr thr asn gly leu val lys thr cys gly gly val tyr gly gln
361/121
CAA ACG TGG CAA TGC CGG GCA GCC TCT GCC CTT GTC GGA TC
gln thr trp gln cys arg ala ala ser ala leu val gly
31/11
91/31
151/51
211/71
271/91
331/111
391/131

```

SEQ ID N° 42C

FIGURE 42C

FEUILLE DE REMPLACEMENT (REGLE 26)

138/185

Séquence codante Rv2622 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq42A:

```

1/1                               31/11
atg gcc aac aaa cgt ggc aat gcc ggg cag cct ctg ccc ttg tcg gat cga gac gac gac
Met ala asn lys arg gly asn ala gly gln pro leu pro leu ser asp arg asp asp asp
61/21                               91/31
cac atg cag ggg cac tgg ctg ctg gcc cgg ctg ggc aag cgg gtg ctg cgt ccc ggc ggc
his met gln gly his trp leu leu ala arg leu gly lys arg val leu arg pro gly gly
121/41                               151/51
gtc gaa ctc acc cgg aca ctg ctg gcc cgc gcc gag gtg acc gac gcc gac gtg ctc gag
val glu leu thr arg thr leu leu ala arg ala glu val thr asp ala asp val leu glu
181/61                               211/71
ctg gca ccg ggc ctg ggc cgc acc gca gcc gaa atc ttg gcc cgc aac ccg cgg tcg tac
leu ala pro gly leu gly arg thr ala ala glu ile leu ala arg asn pro arg ser tyr
241/81                               271/91
gtg ggg gcg gag agc gat ccc aac gcg gcc aac ctg gtc cga cac gtt ctc gcc ggc cgc
val gly ala glu ser asp pro asn ala ala asn leu val arg his val leu ala gly arg
301/101                               331/111
ggc gac gtc cgg gtc acc gac gcg gcc gat acc gga tta tcc gac gcc agc gcc gat gtc
gly asp val arg val thr asp ala ala asp thr gly leu ser asp ala ser ala asp val
361/121                               391/131
gtc atc ggc gag gcg atg ctg acc atg caa ggc aac gcg gct aaa cac acg atc gtc gcc
val ile gly glu ala met leu thr met gln gly asn ala ala lys his thr ile val ala
421/141                               451/151
gag gcg gcg cgg gtg ctg agg ccg ggt ggc cgc tac gcg att cac gaa cta gcg ctg gtg
glu ala ala arg val leu arg pro gly gly arg tyr ala ile his glu leu ala leu val
481/161                               511/171
ccg gac gac gtc gca gag cag gtc cgc acc gac ctg cgg cag tcg ctg gcc cgc gcg ctc
pro asp asp val ala glu gln val arg thr asp leu arg gln ser leu ala arg ala leu
541/181                               571/191
aag gtc aat gcg cgt ccg ctg acc gtt gcg gaa tgg tcg cac ctc tta gcg ggc cat gga
lys val asn ala arg pro leu thr val ala glu trp ser his leu leu ala gly his gly
601/201                               631/211
ctg gtc gtc gaa cac gtt gtc acc gct tcc atg gcg ttg tta caa ccg cga cgg gtg atc
leu val val glu his val val thr ala ser met ala leu leu gln pro arg arg val ile
661/221                               691/231
gct gac gaa ggc ctc ctg ggt gcg ctg cgg ttc gcc gga aac ctg ctc atc cat cgt gcc
ala asp glu gly leu leu gly ala leu arg phe ala gly asn leu leu ile his arg ala
721/241                               751/251
gcg cgt cgg cga gtc ctg ttg atg cgc cac aca ttc cgc agg cat cgt gaa cgc ttg aca
ala arg arg arg val leu leu met arg his thr phe arg arg his arg glu arg leu thr
781/261                               811/271
gcc gtc gcc att gtc gcg cac aaa ccg cac gtc gat tcg tga
ala val ala ile val ala his lys pro his val asp ser OPA

```

SEQ ID N° 42D

FIGURE 42D

139/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv2622

```

1/1                                31/11
taa aaa cct gcg gag gtg ggg tct atg gcc aac aaa cgt ggc aat gcc ggg cag cct ctg
OCH lys pro ala glu val gly ser met ala asn lys arg gly asn ala gly gln pro leu
61/21                                91/31
ccc ttg tcg gat cga gac gac gac cac atg cag ggg cac tgg ctg ctg gcc cgg ctg ggc
pro leu ser asp arg asp asp asp his met gln gly his trp leu leu ala arg leu gly
121/41                                151/51
aag cgg gtg ctg cgt ccc ggc ggc gtc gaa ctc acc cgg aca ctg ctg gcc cgc gcc gag
lys arg val leu arg pro gly gly val glu leu thr arg thr leu leu ala arg ala glu
181/61                                211/71
gtg acc gac gcc gac gtg ctc gag ctg gca ccg ggc ctg ggc cgc acc gca gcc gaa atc
val thr asp ala asp val leu glu leu ala pro gly leu gly arg thr ala ala glu ile
241/81                                271/91
ttg gcc cgc aac ccg cgg tcg tac gtg ggg gcg gag agc gat ccc aac gcg gcc aac ctg
leu ala arg asn pro arg ser tyr val gly ala glu ser asp pro asn ala ala asn leu
301/101                                331/111
gtc cga cac gtt ctc gcc ggc cgc ggc gac gtc cgg gtc acc gac gcg gcc gat acc gga
val arg his val leu ala gly arg gly asp val arg val thr asp ala ala asp thr gly
361/121                                391/131
tta tcc gac gcc agc gcc gat gtc gtc atc ggc gag gcg atg ctg acc atg caa ggc aac
leu ser asp ala ser ala asp val val ile gly glu ala met leu thr met gln gly asn
421/141                                451/151
gcg gct aaa cac acg atc gtc gcc gag gcg gcg cgg gtg ctg agg ccg ggt ggc cgc tac
ala ala lys his thr ile val ala glu ala ala arg val leu arg pro gly gly arg tyr
481/161                                511/171
gcg att cac gaa cta gcg ctg gtg ccg gac gac gtc gca gag cag gtc cgc acc gac ctg
ala ile his glu leu ala leu val pro asp asp val ala glu gln val arg thr asp leu
541/181                                571/191
cgg cag tcg ctg gcc cgc gcg ctc aag gtc aat gcg cgt ccg ctg acc gtt gcg gaa tgg
arg gln ser leu ala arg ala leu lys val asn ala arg pro leu thr val ala glu trp
601/201                                631/211
tcg cac ctc tta gcg ggc cat gga ctg gtc gtc gaa cac gtt gtc acc gct tcc atg gcg
ser his leu leu ala gly his gly leu val val glu his val val thr ala ser met ala
661/221                                691/231
ttg tta caa ccg cga cgg gtg atc gct gac gaa ggc ctc ctg ggt gcg ctg cgg ttc gcc
leu leu gln pro arg arg val ile ala asp glu gly leu leu gly ala leu arg phe ala
721/241                                751/251
gga aac ctg ctc atc cat cgt gcc gcg cgt cgg cga gtc ctg ttg atg cgc cac aca ttc
gly asn leu leu ile his arg ala ala arg arg arg val leu leu met arg his thr phe
781/261                                811/271
cgc agg cat cgt gaa cgc ttg aca gcc gtc gcc att gtc gcg cac aaa ccg cac gtc gat
arg arg his arg glu arg leu thr ala val ala ile val ala his lys pro his val asp
841/281
tcg tga
ser OPA

```

SEQ ID N° 42F

FIGURE 42F

FEUILLE DE REMPLACEMENT (REGLE 26)

140/185

```

1/1                                31/11
atc gcg cgt gac atc gat gac cag ggt cgg ctg tgt ctg gac gtc ggc ggt cga acg gta
ile ala arg asp ile asp asp gln gly arg leu cys leu asp val gly gly arg thr val
61/21                                91/31
gtt gtt tca gcg ggc gac gtg gtg cat ttg cgt taa ctc gcg cgg agc tgg cgt ccc caa
val val ser ala gly asp val val his leu arg OCH leu ala arg ser trp arg pro gln
121/41                                151/51
aag att aag gtc gcg ggc atg agc tat ccg gag aat gtc ctg gcc gct ggc gag cag gtc
lys ile lys val ala gly met ser tyr pro glu asn val leu ala ala gly glu gln val
181/61                                211/71
gtt ctg cac cgc cat ccg cac tgg aat cgc tta atc tgg ccc gtc gtg gtg ctg gtc ttg
val leu his arg his pro his trp asn arg leu ile trp pro val val val leu val leu
241/81                                271/91
ctg acc ggg ttg gcg gcg ttc ggg tcc gga ttc gtc aac tcg aca cct tgg cag cag atc
leu thr gly leu ala ala phe gly ser gly phe val asn ser thr pro trp gln gln ile

```

SEQ ID N° 43A

FIGURE 43A

```

1/1                                31/11
tcg cgc gtg aca tcg atg acc agg gtc ggc tgt gtc tgg acg tcg gcg gtc gaa cgg tag
ser arg val thr ser met thr arg val gly cys val trp thr ser ala val glu arg AMB
61/21                                91/31
ttg ttt cag .cgg gcg acg tgg tgc att tgc gtt aac tcg cgc gga gct ggc gtc ccc aaa
leu phe gln arg ala thr trp cys ile cys val asn ser arg gly ala gly val pro lys
121/41                                151/51
aga tta agg tcg cgg gca tga gct atc ccg aga atg tcc tgg ccg ctg gcg agc agg tcg
arg leu arg ser arg ala OPA ala ile arg arg met ser trp pro leu ala ser arg ser
181/61                                211/71
ttc tgc acc gcc atc cgc act gga atc gct taa tct ggc ccg tcg tgg tgc tgg tct tgc
phe cys thr ala ile arg thr gly ile ala OCH ser gly pro ser trp cys trp ser cys
241/81                                271/91
tga ccg ggt tgg cgg cgt tcg ggt ccg gat tcg tca act cga cac ctt ggc agc aga tc
OPA pro gly trp arg arg ser gly pro asp ser ser thr arg his leu gly ser arg

```

SEQ ID N° 43B

FIGURE 43B

141/185

```

1/1                               31/11
cgc gcg tga cat cga tga cca ggg tcg gct gtg tct gga cgt cgg cgg tcg aac ggt agt
arg ala OPA his arg OPA pro gly ser ala val ser gly arg arg arg ser asn gly ser
61/21                               91/31
tgt ttc agc ggg cga cgt ggt gca ttt gcg tta act cgc gcg gag ctg gcg tcc cca aaa
cys phe ser gly arg arg gly ala phe ala leu thr arg ala glu leu ala ser pro lys
121/41                               151/51
gat taa ggt cgc ggg cat gag cta tcc gga gaa tgt cct ggc cgc tgg cga gca ggt cgt
asp OCH gly arg gly his glu leu ser gly glu cys pro gly arg trp arg ala gly arg
181/61                               211/71
tct gca ccg cca tcc gca ctg gaa tcg ctt aat ctg gcc cgt cgt ggt gct ggt ctt gct
ser ala pro pro ser ala leu glu ser leu asn leu ala arg arg gly ala gly leu ala
241/81                               271/91
gac cgg gtt ggc ggc gtt cgg gtc cgg att cgt caa ctc gac acc ttg gca gca gat c
asp arg val gly gly val arg val arg ile arg gln leu asp thr leu ala ala asp

```

SEQ ID N° 43C

FIGURE 43C

Séquence codante Rv3278c prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq43A:

```

1/1                               31/11
atg agc tat ccg gag aat gtc ctg gcc gct ggc gag cag gtc gtt ctg cac cgc cat ccg
Met ser tyr pro glu asn val leu ala ala gly glu gln val val leu his arg his pro
61/21                               91/31
cac tgg aat cgc tta atc tgg ccc gtc gtg gtg ctg gtc ttg ctg acc ggg ttg gcg gcg
his trp asn arg leu ile trp pro val val val leu val leu leu thr gly leu ala ala
121/41                               151/51
ttc ggg tcc gga ttc gtc aac tcg aca cct tgg cag cag atc gct aag aac gtg att cac
phe gly ser gly phe val asn ser thr pro trp gln gln ile ala lys asn val ile his
181/61                               211/71
gcg gtc atc tgg ggg atc tgg ttg gtg atc gtc gcc tgg ctc acg ctg tgg cca ttc ctg
ala val ile trp gly ile trp leu val ile val gly trp leu thr leu trp pro phe leu
241/81                               271/91
agc tgg ctg acc aca cat ttc gtg gtg acc aac cgg cgg gtg atg ttc cgg cat ggt gtg
ser trp leu thr thr his phe val val thr asn arg arg val met phe arg his gly val
301/101                               331/111
ctg acc cgc agc ggg atc gac ata ccg cta gca cgg atc aac agc gtg gag ttc cgg gac
leu thr arg ser gly ile asp ile pro leu ala arg ile asn ser val glu phe arg asp
361/121                               391/131
cgg atc ttc gag cgg att ttt cgc acc ggg acg ttg att atc gag tcc gcg tca caa gat
arg ile phe glu arg ile phe arg thr gly thr leu ile ile glu ser ala ser gln asp
421/141                               451/151
ccg ctc gag ttc tac aac att ccg cgc ctg cgg gag gtg cat gcg ttg ctg tat cac gag
pro leu glu phe tyr asn ile pro arg leu arg glu val his ala leu leu tyr his glu
481/161                               511/171
gtt ttc gac acc ctg ggc tcc gac gag tcg ccc agc tga
val phe asp thr leu gly ser asp glu ser pro ser OPA

```

SEQ ID N° 43D

FIGURE 43D

FEUILLE DE REMPLACEMENT (REGLE 26)

142/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv3278c

1/1	31/11
taa ctc gcg cgg agc tgg cgt ccc caa aag	att aag gtc gcg ggc atg agc tat ccg gag
OCH leu ala arg ser trp arg pro gln lys	ile lys val ala gly met ser tyr pro glu
61/21	91/31
aat gtc ctg gcc gct ggc gag cag gtc gtt	ctg cac cgc cat ccg cac tgg aat cgc tta
asn val leu ala ala gly glu gln val val	leu his arg his pro his trp asn arg leu
121/41	151/51
atc tgg ccc gtc gtg gtg ctg gtc ttg ctg	acc ggg ttg gcg gcg ttc ggg tcc gga ttc
ile trp pro val val val leu val leu leu	thr gly leu ala ala phe gly ser gly phe
181/61	211/71
gtc aac tcg aca cct tgg cag cag atc gct	aag aac gtg att cac gcg gtc atc tgg ggg
val asn ser thr pro trp gln gln ile ala	lys asn val ile his ala val ile trp gly
241/81	271/91
atc tgg ttg gtg atc gtc ggc tgg ctc acg	ctg tgg cca ttc ctg agc tgg ctg acc aca
ile trp leu val ile val gly trp leu thr	leu trp pro phe leu ser trp leu thr thr
301/101	331/111
cat ttc gtg gtg acc aac cgg cgg gtg atg	ttc cgg cat ggt gtg ctg acc cgc agc ggg
his phe val val thr asn arg arg val met	phe arg his gly val leu thr arg ser gly
361/121	391/131
atc gac ata ccg cta gca cgg atc aac agc	gtg gag ttc cgg gac cgg atc ttc gag cgg
ile asp ile pro leu ala arg ile asn ser	val glu phe arg asp arg ile phe glu arg
421/141	451/151
att ttt cgc acc ggg acg ttg att atc gag	tcc gcg tca caa gat ccg ctc gag ttc tac
ile phe arg thr gly thr leu ile ile glu	ser ala ser gln asp pro leu glu phe tyr
481/161	511/171
aac att ccg cgc ctg cgg gag gtg cat gcg	ttg ctg tat cac gag gtt ttc gac acc ctg
asn ile pro arg leu arg glu val his ala	leu leu tyr his glu val phe asp thr leu
541/181	
ggc tcc gac gag tcg ccc agc tga	
gly ser asp glu ser pro ser OPA	

SEQ ID N° 43F

FIGURE 43F

143/185

```

1/1                               31/11
gcc aag atg gat gtc tac caa cgc acc gcc gcc ggc tgg cag ccg ctc aag acc ggt atc
ala lys met asp val tyr gln arg thr ala ala gly trp gln pro leu lys thr gly ile
61/21                               91/31
acc acc cat atc ggt tcg gcg ggc atg gcg ccg gaa gcc aag agc gga tat ccg gcc act
thr thr his ile gly ser ala gly met ala pro glu ala lys ser gly tyr pro ala thr
121/41                               151/51
ccg atg ggg gtt tac agc ctg gac tcc gct ttt ggc acc gcg ccg aat ccc ggt ggc ggg
pro met gly val tyr ser leu asp ser ala phe gly thr ala pro asn pro gly gly gly
181/61                               211/71
ttg ccg tat acc caa gtc gga ccc aat cac tgg tgg agt ggc gac gac aat agc ccc acc
leu pro tyr thr gln val gly pro asn his trp trp ser gly asp asp asn ser pro thr
241/81                               271/91
ttt aac tcc atg cag gtc tgt cag aag tcc cag tgc ccg ttc agc acg gcc gac agc gag
phe asn ser met gln val cys gln lys ser gln cys pro phe ser thr ala asp ser glu
301/101                               331/111
aac ctg caa atc ccg cag tac aag cat tcg gtc gtg atg ggc gtc aac aag gcc aag gtc
asn leu gln ile pro gln tyr lys his ser val val met gly val asn lys ala lys val
361/121                               391/131
cca ggc aaa ggc tcc gcg ttc ttc ttt cac acc acc gac ggc ggg ccc acc gcg ggt tgt
pro gly lys gly ser ala phe phe phe his thr thr asp gly gly pro thr ala gly cys
421/141
gtg gcg atc
val ala ile

```

SEQ ID N° 44A

FIGURE 44A

```

1/1                               31/11
cca aga tgg atg tct acc aac gca ccg ccg ccg gct ggc agc cgc tca aga ccg gta tca
pro arg trp met ser thr asn ala pro pro pro ala gly ser arg ser arg pro val ser
61/21                               91/31
cca ccc ata tcg gtt cgg cgg gca tgg cgc ccg aag cca aga gcg gat atc ccg cca ctc
pro pro ile ser val arg arg ala trp arg arg lys pro arg ala asp ile arg pro leu
121/41                               151/51
cga tgg ggg ttt aca gcc tgg act ccg ctt ttg gca ccg cgc cga atc ccg gtg gcg ggt
arg trp gly phe thr ala trp thr pro leu leu ala pro arg arg ile pro val ala gly
181/61                               211/71
tgc cgt ata ccc aag tcg gac cca atc act ggt gga gtg gcg acg aca ata gcc cca cct
cys arg ile pro lys ser asp pro ile thr gly gly val ala thr thr ile ala pro pro
241/81                               271/91
tta act cca tgc agg tct gtc aga agt ccc agt gcc cgt tca gca ccg ccg aca gcg aga
leu thr pro cys arg ser val arg ser pro ser ala arg ser ala arg pro thr ala arg
301/101                               331/111
acc tgc aaa tcc cgc agt aca agc att cgg tcg tga tgg gcg tca aca agg cca agg tcc
thr cys lys ser arg ser thr ser ile arg ser opa trp ala ser thr arg pro arg ser
361/121                               391/131
cag gca aag gct ccg cgt tct tct ttc aca cca ccg acg gcg ggc cca ccg ccg gtt gtg
gln ala lys ala pro arg ser ser phe thr pro pro thr ala gly pro pro arg val val
421/141
tgg cga tc
trp arg

```

SEQ ID N° 44B

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 44B

144/185

1/1
 caa gat gga tgt cta cca acg cac cgc cgc cgg ctg gca gcc gct caa gac cgg tat cac
 gln asp gly cys leu pro thr his arg arg arg leu ala ala ala gln asp arg tyr his
 61/21
 cac cca tat cgg ttc ggc ggg cat ggc gcc gga agc caa gag cgg ata tcc ggc cac tcc
 his pro tyr arg phe gly gly his gly ala gly ser gln glu arg ile ser gly his ser
 121/41
 gat ggg ggt tta cag cct gga ctc cgc ttt tgg cac cgc gcc gaa tcc cgg tgg cgg gtt
 asp gly gly leu gln pro gly leu arg phe trp his arg ala glu ser arg trp arg val
 181/61
 gcc gta tac cca agt cgg acc caa tca ctg gtg gag tgg cga cga caa tag ccc cac ctt
 ala val tyr pro ser arg thr gln ser leu val glu trp arg arg gln AMB pro his leu
 241/81
 taa ctc cat gca ggt ctg tca gaa gtc cca gtg ccc gtt cag cac ggc cga cag cga gaa
 OCH leu his ala gly leu ser glu val pro val pro val gln his gly arg gln arg glu
 301/101
 cct gca aat ccc gca gta caa gca ttc ggt cgt gat ggg cgt caa caa ggc caa ggt ccc
 pro ala asn pro ala val gln ala phe gly arg asp gly arg gln gln gly gln gly pro
 361/121
 agg caa agg ctc cgc gtt ctt ctt tca cac cac cga cgg cgg gcc cac cgc ggg ttg tgt
 arg gln arg leu arg val leu leu ser his his arg arg arg ala his arg gly leu cys
 421/141
 ggc gat c
 gly asp

SEQ ID N° 44C

FIGURE 44C

145/185

Séquence codante Rv0309 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant Séq44A:

```

1/1                               31/11
atg agc cga ctc cta gct ttg ctg tgc gct gcg gta tgc acg ggc tgc gtt gct gtg gtt
Met ser arg leu leu ala leu leu cys ala ala val cys thr gly cys val ala val val
61/21                               91/31
ctc gcg cca gtg agc ctg gcc gtc gtc aac ccg tgg ttc gcg aac tcg gtc ggc aat gcc
leu ala pro val ser leu ala val val asn pro trp phe ala asn ser val gly asn ala
121/41                               151/51
act cag gtg gtt tcg gtg gtg gga acc ggc ggt tcg acg gcc aag atg gat gtc tac caa
thr gln val val ser val val gly thr gly gly ser thr ala lys met asp val tyr gln
181/61                               211/71
cgc acc gcc gcc ggc tgg cag ccg ctc aag acc ggt atc acc acc cat atc ggt tcg gcg
arg thr ala ala gly trp gln pro leu lys thr gly ile thr thr his ile gly ser ala
241/81                               271/91
ggc atg gcg ccg gaa gcc aag agc gga tat ccg gcc act ccg atg ggg gtt tac agc ctg
gly met ala pro glu ala lys ser gly tyr pro ala thr pro met gly val tyr ser leu
301/101                               331/111
gac tcc gct ttt ggc acc gcg ccg aat ccc ggt ggc ggg ttg ccg tat acc caa gtc gga
asp ser ala phe gly thr ala pro asn pro gly gly gly leu pro tyr thr gln val gly
361/121                               391/131
ccc aat cac tgg tgg agt ggc gac gac aat agc ccc acc ttt aac tcc atg cag gtc tgt
pro asn his trp trp ser gly asp asp asn ser pro thr phe asn ser met gln val cys
421/141                               451/151
cag aag tcc cag tgc ccg ttc agc acg gcc gac agc gag aac ctg caa atc ccg cag tac
gln lys ser gln cys pro phe ser thr ala asp ser glu asn leu gln ile pro gln tyr
481/161                               511/171
aag cat tcg gtc gtg atg ggc gtc aac aag gcc aag gtc cca ggc aaa ggc tcc gcg ttc
lys his ser val val met gly val asn lys ala lys val pro gly lys gly ser ala phe
541/181                               571/191
ttc ttt cac acc acc gac ggc ggg ccc acc gcg ggt tgt gtg gcg atc gac gat gcc acg
phe phe his thr thr asp gly gly pro thr ala gly cys val ala ile asp asp ala thr
601/201                               631/211
ctg gtg cag atc atc cgt tgg ctg cgg cct ggt gcg gtg atc gcg atc gcc aag taa
leu val gln ile ile arg trp leu arg pro gly ala val ile ala ile ala lys OCH

```

SEQ ID N° 44D

FIGURE 44D

146/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0309

```

1/1                                31/11
tga gcg atg agc cga ctc cta gct ttg ctg tgc gct gcg gta tgc acg ggc tgc gtt gct
OPA ala met ser arg leu leu ala leu leu cys ala ala val cys thr gly cys val ala
61/21                                91/31
gtg gtt ctc gcg cca gtg agc ctg gcc gtc gtc aac ccg tgg ttc gcg aac tcg gtc ggc
val val leu ala pro val ser leu ala val val asn pro trp phe ala asn ser val gly
121/41                                151/51
aat gcc act cag gtg gtt tcg gtg gtg gga acc ggc ggt tcg acg gcc aag atg gat gtc
asn ala thr gln val val ser val val gly thr gly gly ser thr ala lys met asp val
181/61                                211/71
tac caa cgc acc gcc gcc gcc tgg cag ccg ctc aag acc ggt atc acc acc cat atc ggt
tyr gln arg thr ala ala gly trp gln pro leu lys thr gly ile thr thr his ile gly
241/81                                271/91
tcg gcg ggc atg gcg ccg gaa gcc aag agc gga tat ccg gcc act ccg atg ggg gtt tac
ser ala gly met ala pro glu ala lys ser gly tyr pro ala thr pro met gly val tyr
301/101                                331/111
agc ctg gac tcc gct ttt ggc acc gcg ccg aat ccc ggt ggc ggg ttg ccg tat acc caa
ser leu asp ser ala phe gly thr ala pro asn pro gly gly gly leu pro tyr thr gln
361/121                                391/131
gtc gga ccc aat cac tgg tgg agt ggc gac gac aat agc ccc acc ttt aac tcc atg cag
val gly pro asn his trp trp ser gly asp asp asn ser pro thr phe asn ser met gln
421/141                                451/151
gtc tgt cag aag tcc cag tgc ccg ttc agc acg gcc gac agc gag aac ctg caa atc ccg
val cys gln lys ser gln cys pro phe ser thr ala asp ser glu asn leu gln ile pro
481/161                                511/171
cag tac aag cat tcg gtc gtg atg ggc gtc aac aag gcc aag gtc cca ggc aaa ggc tcc
gln tyr lys his ser val val met gly val asn lys ala lys val pro gly lys gly ser
541/181                                571/191
gcg ttc ttc ttt cac acc acc gac ggc ggg ccc acc gcg ggt tgt gtg gcg atc gac gat
ala phe phe phe his thr thr asp gly gly pro thr ala gly cys val ala ile asp asp
601/201                                631/211
gcc acg ctg gtg cag atc atc cgt tgg ctg cgg cct ggt gcg gtg atc gcg atc gcc aag
ala thr leu val gln ile ile arg trp leu arg pro gly ala val ile ala ile ala lys
661/221
taa
OCH

```

SEQ ID N° 44F

FIGURE 44F

147/185

Fragment cloné en fusion avec phoA

1/1	31/11
gat ctc ccc gga cac cag gtc atc cgg cga	gat ggt gat cga ggc tcg gac ccg cag gca
asp leu pro gly his gln val ile arg arg	asp gly asp arg gly ser asp pro gln ala
61/21	91/31
tcc ggt agc cag agg cac cag cat cag caa	cat cgc gat ggc cag cat gcc gcg ccg tcg
ser gly ser gln arg his gln his gln gln	his arg asp gly gln his ala ala pro ser
121/41	151/51
ggt cct tgc cac tcg cga tcc ttg gga tga	cgg tgg ggc ata gct agc gcg cac cag gtc
gly pro cys his ser arg ser leu gly OPA	arg trp gly ile ala ser ala his gln val
181/61	211/71
atc gtg cca gac cgg gca tgc cgc gtc ggc	aag ctg tcg ggc gcg ggt tag agc ggt agc
ile val pro asp arg ala cys arg val gly	lys leu ser gly ala gly AMB ser gly ser
241/81	271/91
gtg cga ccc agg atg gcg aat gct cgg ggg	tca ccg gcg aag tgg tag ccg ccg atg atg
val arg pro arg met ala asn ala arg gly	ser pro ala lys trp AMB pro arg met met
301/101	331/111
tcg gtg aag ccc aac cgg cgg tac aac cgc	cac gcc cga ttg tcc tca ccg ttg gtc tcc
ser val lys pro asn arg arg tyr asn arg	his ala arg leu ser ser pro leu val ser
361/121	391/131
ggt gtg gag agc agg acg ttg tcc tcg tcg	cga ccg gct agc agt ccg ccg gcc aac gcc
gly val glu ser arg thr leu ser ser ser	arg pro ala ser ser arg arg ala asn ala
421/141	451/151
tcc ccg agg cca cgg cct tga gcg cgg gga	agg atg tgc aat tca gtc aac tcg aag tag
ser pro arg pro arg pro OPA ala arg gly	arg met cys asn ser val asn ser lys AMB
481/161	511/171
ctg gtc atc agt cgg gcg atc gct agg cgc	gga aag ccg ctg cgt tgc aag ccc agt acc
leu val ile ser arg ala ile ala arg arg	gly lys pro leu arg cys lys pro ser thr
541/181	571/191
acc tgc tgt tgc cac cac tgg ccg ggc gcc	ccg gga tag ccg tac gcc act ccg agc att
thr cys cys cys his his trp pro gly ala	pro gly AMB pro tyr ala thr pro ser ile
601/201	631/211
ggc gcg ttg ctc agt tcg gcg gcc gac ggc	agc gcc gtg gtg tcg gcg gcc tcg gcc tgt
gly ala leu leu ser ser ala ala asp gly	ser ala val val ser ala ala ser ala cys
661/221	691/231
tcg gct gcc gtt acc tcg acg gcc gcg acc	gcc tgc cag ccg cgc cgc ccg atg tgc tcc
ser ala ala val thr ser thr ala ala thr	ala cys gln pro arg arg arg met cys ser
721/241	751/251
agc cac att ggg gcg cgc aaa gtc tcg gtg	ccc ctg ggg tag cgc atc gcg tcg aca tac
ser his ile gly ala arg lys val ser val	pro leu gly AMB arg ile ala ser thr tyr
781/261	811/271
acc gtc agg gca tca ccg agg cgg cgc tcc	ata tcg ctg ggc ggc aga tcg atg agg aat
thr val arg ala ser pro arg arg arg ser	ile ser leu gly gly arg ser met arg asn
841/281	871/291
atc gcc aac gcg cgg tgt cct cct cat gtg	atg aac cga tgc gtg ctt gcg cac cag tat
ile ala asn ala arg cys pro pro his val	met asn arg cys val leu ala his gln tyr
901/301	931/311
cgg aca agc cga tga ggc cgc ccg cgc tgg	acg ggg ctt gta gcg tat ggc cgt ttc cgc
arg thr ser arg OPA gly arg pro arg trp	thr gly leu val ala tyr gly arg phe arg

SEQ ID N° 45ZA

FIGURE 45ZA

FEUILLE DE REMPLACEMENT (REGLE 26)

148/185

961/321	991/331
tca gct cgt cgc tgc ggc gcc gcc ggg ata	gaa tcg ccc gcg aac cag tgg tac ggc gca
ser ala arg arg cys gly ala ala gly ile	glu ser pro ala asn gln trp tyr gly ala
1021/341	1051/351
gat tga cct cgt atc atc tga gtt agt tgc	ccg cgc aat ggg cat ccg cgt gtt atc ggt
asp opa pro arg ile ile opa val ser cys	pro arg asn gly his pro arg val ile gly
1081/361	1111/371
att acg tga cag tct gtc ggc aag gag gga	cgc atg cca ctc tcc gat cat gag cag cgg
ile thr opa gln ser val gly lys glu gly	arg met pro leu ser asp his glu gln arg
1141/381	1171/391
atg ctt gac cag atc gag agc gct ctc tac	gcc gaa gat ccc aag ttc gca tcg agt gtc
met leu asp gln ile glu ser ala leu tyr	ala glu asp pro lys phe ala ser ser val
1201/401	1231/411
cgt ggc ggg ggc ttc cgc gca ccg acc gcg	cgg cgg cgc ctg cag ggc gcg gcg ttg ttc
arg gly gly phe arg ala pro thr ala	arg arg arg leu gln gly ala ala leu phe
1261/421	1291/431
atc atc ggt ctg ggg atg ttg gtt tcc ggc	gtg gcg ttc aaa gag acc atg atc gga agt
ile ile gly leu gly met leu val ser gly	val ala phe lys glu thr met ile gly ser
1321/441	1351/451
ttc ccg ata ctc agc gtt ttc ggt ttt gtc	gtg atg ttc ggt ggt gtg gtg tat gcc atc
phe pro ile leu ser val phe gly phe val	val met phe gly gly val val tyr ala ile
1381/461	1411/471
acc ggt cct cgg ttg tcc ggc agg atg gat	cgt ggc gga tcg gct gct ggg gct tcg cgc
thr gly pro arg leu ser gly arg met asp	arg gly gly ser ala ala gly ala ser arg
1441/481	1471/491
cag cgt cgt acc aag ggg gcc ggg ggc tca	ttc acc agc cgt atg gaa gat c
gln arg arg thr lys gly ala gly gly ser	phe thr ser arg met glu asp

SEQ ID N° 452A (suite)

FIGURE 452A (suite)

149/185

fragment seq452A en décalage moins 1 pour la phase de lecture

```

1/1                                31/11
atc tcc ccg gac acc agg tca tcc ggc gag atg gtg atc gag gct cgg acc cgc agg cat
ile ser pro asp thr arg ser ser gly glu met val ile glu ala arg thr arg arg his
61/21                                91/31
ccg gta gcc aga ggc acc agc atc agc aac atc gcg atg gcc agc atg ccg cgc cgt cgg
pro val ala arg gly thr ser ile ser asn ile ala met ala ser met pro arg arg arg
121/41                                151/51
gtc ctt gcc act cgc gat cct tgg gat gac ggt ggg gca tag cta gcg cgc acc agg tca
val leu ala thr arg asp pro trp asp asp gly gly ala AMB leu ala arg thr arg ser
181/61                                211/71
tcg tgc cag acc ggg cat gcc gcg tcg gca agc tgt cgg gcg cgg gtt aga gcg gta gcg
ser cys gln thr gly his ala ala ser ala ser cys arg ala arg val arg ala val ala
241/81                                271/91
tgc gac cca gga tgg cga atg ctc ggg ggt cac cgg cga agt ggt agc cgc gga tga tgt
cys asp pro gly trp arg met leu gly gly his arg arg ser gly ser arg gly OPA cys
301/101                                331/111
cgg tga agc cca acc ggc ggt aca acc gcc acg ccc gat tgt cct cac cgt tgg tct ccg
arg OPA ser pro thr gly gly thr thr ala thr pro asp cys pro his arg trp ser pro
361/121                                391/131
gtg tgg aga gca gga cgt tgt cct cgt cgc gac cgg cta gca gtc ggc ggg cca acg cct
val trp arg ala gly arg cys pro arg arg asp arg leu ala val gly gly pro thr pro
421/141                                451/151
ccc cga ggc cac ggc ctt gag cgc ggg gaa gga tgt gca att cag tca act cga agt agc
pro arg gly his gly leu glu arg gly glu gly cys ala ile gln ser thr arg ser ser
481/161                                511/171
tgg tca tca gtc ggg cga tcg cta ggc gcg gaa agc cgc tgc gtt gca agc cca gta cca
trp ser ser val gly arg ser leu gly ala glu ser arg cys val ala ser pro val pro
541/181                                571/191
cct gct gtt gcc acc act ggc cgg gcg ccc cgg gat agc cgt acg cca ctc cga gca ttg
pro ala val ala thr thr gly arg ala pro arg asp ser arg thr pro leu arg ala leu
601/201                                631/211
gcg cgt tgc tca gtt cgg cgg ccg acg gca gcg ccg tgg tgt cgg cgg cct cgg cct gtt
ala arg cys ser val arg arg pro thr ala ala pro trp cys arg arg pro arg pro val
661/221                                691/231
cgg ctg ccg tta cct cga cgg ccg cga ccg cct gcc agc cgc gcc gcc gga tgt gct cca
arg leu pro leu pro arg arg pro arg pro pro ala ser arg ala ala gly cys ala pro
721/241                                751/251
gcc aca ttg ggg cgc gca aag tct cgg tgc ccc tgg ggt agc gca tcg cgt cga cat aca
ala thr leu gly arg ala lys ser arg cys pro trp gly ser ala ser arg arg his thr
781/261                                811/271
ccg tca ggg cat cac cga ggc ggc gct cca tat cgc tgg gcg gca gat cga tga gga ata
pro ser gly his his arg gly gly ala pro tyr arg trp ala ala asp arg OPA gly ile
841/281                                871/291
tcg cca acg cgc ggt gtc ctc ctc atg tga tga acc gat gcg tgc ttg cgc acc agt atc
ser pro thr arg gly val leu leu met OPA OPA thr asp ala cys leu arg thr ser ile
901/301                                931/311
gga caa gcc gat gag gcc gcc cgc gct gga cgg ggc ttg tag cgt atg gcc gtt tcc gct
gly gln ala asp glu ala ala arg ala gly arg gly leu AMB arg met ala val ser ala
961/321                                991/331
cag ctc gtc gct gcg gcg ccg ccg gga tag aat cgc ccg cga acc agt ggt acg gcg cag
gln leu val ala ala ala pro pro gly AMB asn arg pro arg thr ser gly thr ala gln

```

SEQ ID N° 452B

FIGURE 452B

FEUILLE DE REMPLACEMENT (REGLE 26)

150/185

1021/341

att gac ctc gta tca tct gag tta gtt gcc cgc gca atg ggc atc cgc gtg tta tcg gta
 ile asp leu val ser ser glu leu val ala arg ala met gly ile arg val leu ser val

1081/361

tta cgt gac agt ctg tcg gca agg agg gac gca tgc cac tct ccg atc atg agc agc gga
 leu arg asp ser leu ser ala arg arg asp ala cys his ser pro ile met ser ser gly

1141/381

tgc ttg acc aga tcg aga gcg ctc tct acg ccg aag atc cca agt tcg cat cga gtg tcc
 cys leu thr arg ser arg ala leu ser thr pro lys ile pro ser ser his arg val ser

1201/401

gtg gcg ggg gct tcc gcg cac cga ccg cgc ggc ggc gcc tgc agg gcg cgg cgt tgt tca
 val ala gly ala ser ala his arg pro arg gly gly ala cys arg ala arg arg cys ser

1261/421

tca tcg gtc tgg gga tgt tgg ttt ccg gcg tgg cgt tca aag aga cca tga tcg gaa gtt
 ser ser val trp gly cys trp phe pro ala trp arg ser lys arg pro OPA ser glu val

1321/441

tcc cga tac tca gcg ttt tcg gtt ttg tcg tga tgt tcg gtg gtg tgg tgt atg cca tca
 ser arg tyr ser ala phe ser val leu ser OPA cys ser val val trp cys met pro ser

1381/461

ccg gtc ctc ggt tgt ccg gca gga tgg atc gtg gcg gat ccg ctg ctg ggg ctt cgc gcc
 pro val leu gly cys pro ala gly trp ile val ala asp arg leu leu gly leu arg ala

1441/481

agc gtc gta cca agg ggg ccg ggg gct cat tca cca gcc gta tgg aag atc
 ser val val pro arg gly pro gly ala his ser pro ala val trp lys ile

1051/351

1111/371

1171/391

1231/411

1291/431

1351/451

1411/471

1471/491

SEQ ID N° 45ZB (suite)

FIGURE 45ZB (suite)

151/185

fragment seq452A en décalage moins 2 pour la phase de lecture

```

1/1                                31/11
tct ccc cgg aca cca ggt cat ccg gcg aga tgg tga tcg agg ctc gga ccc gca ggc atc
ser pro arg thr pro gly his pro ala arg trp OPA ser arg leu gly pro ala gly ile
61/21                                91/31
cgg tag cca gag gca cca gca tca gca aca tcg cga tgg cca gca tgc cgc gcc gtc ggg
arg AMB pro glu ala pro ala ser ala thr ser arg trp pro ala cys arg ala val gly
121/41                                151/51
tcc ttg cca ctc gcg atc ctt ggg atg acg gtg ggg cat agc tag cgc gca cca ggt cat
ser leu pro leu ala ile leu gly met thr val gly his ser AMB arg ala pro gly his
181/61                                211/71
cgt gcc aga ccg ggc atg ccg cgt cgg caa gct gtc ggg cgc ggg tta gag cgg tag cgt
arg ala arg pro gly met pro arg arg gln ala val gly arg gly leu glu arg AMB arg
241/81                                271/91
gcg acc cag gat ggc gaa tgc tcg ggg gtc acc ggc gaa gtg gta gcc gcg gat gat gtc
ala thr gln asp gly glu cys ser gly val thr gly glu val val ala ala asp asp val
301/101                                331/111
ggt gaa gcc caa ccg gcg gta caa ccg cca cgc ccg att gtc ctc acc gtt ggt ctc cgg
gly glu ala gln pro ala val gln pro pro arg pro ile val leu thr val gly leu arg
361/121                                391/131
tgt gga gag cag gac gtt gtc ctc gtc gcg acc ggc tag cag tcg gcg ggc caa cgc ctc
cys gly glu gln asp val val leu val ala thr gly AMB gln ser ala gly gln arg leu
421/141                                451/151
ccc gag gcc acg gcc ttg agc gcg ggg aag gat gtg caa ttc agt caa ctc gaa gta gct
pro glu ala thr ala leu ser ala gly lys asp val gln phe ser gln leu glu val ala
481/161                                511/171
ggt cat cag tcg ggc gat cgc tag gcg cgg aaa gcc gct gcg ttg caa gcc cag tac cac
gly his gln ser gly asp arg AMB ala arg lys ala ala ala leu gln ala gln tyr his
541/181                                571/191
ctg ctg ttg cca cca ctg gcc ggg cgc ccc ggg ata gcc gta cgc cac tcc gag cat tgg
leu leu leu pro pro leu ala gly arg pro gly ile ala val arg his ser glu his trp
601/201                                631/211
cgc gtt gct cag ttc ggc ggc cga cgg cag cgc cgt ggt gtc ggc ggc ctc ggc ctg ttc
arg val ala gln phe gly gly arg arg gln arg arg gly val gly gly leu gly leu phe
661/221                                691/231
ggc tgc cgt tac ctc gac ggc cgc gac cgc ctg cca gcc gcg ccg ccg gat gtg ctc cag
gly cys arg tyr leu asp gly arg asp arg leu pro ala ala pro pro asp val leu gln
721/241                                751/251
cca cat tgg ggc gcg caa agt ctc ggt gcc cct ggg gta gcg cat cgc gtc gac ata cac
pro his trp gly ala gln ser leu gly ala pro gly val ala his arg val asp ile his
781/261                                811/271
cgt cag ggc atc acc gag gcg gcg ctc cat atc gct ggg cgg cag atc gat gag gaa tat
arg gln gly ile thr glu ala ala leu his ile ala gly arg gln ile asp glu glu tyr
841/281                                871/291
cgc caa cgc gcg gtg tcc tcc tca tgt gat gaa ccg atg cgt gct tgc gca cca gta tcg
arg gln arg ala val ser ser ser cys asp glu pro met arg ala cys ala pro val ser
901/301                                931/311
gac aag ccg atg agg ccg ccc gcg ctg gac ggg gct tgt agc gta tgg ccg ttt ccg ctc
asp lys pro met arg pro pro ala leu asp gly ala cys ser val trp pro phe pro leu

```

SEQ ID N° 452C

FIGURE 452C

FEUILLE DE REMPLACEMENT (REGLE 26)

152/185

961/321
 agc tcg tcg ctg cgg cgc cgc cgg gat aga atc gcc cgc gaa cca gtg gta cgg cgc aga
 ser ser ser leu arg arg arg arg asp arg ile ala arg glu pro val val arg arg arg
 1021/341
 ttg acc tcg tat cat ctg agt tag ttg ccc gcg caa tgg gca tcc gcg tgt tat cgg tat
 leu thr ser tyr his leu ser AMB leu pro ala gln trp ala ser ala cys tyr arg tyr
 1081/361
 tac gtg aca gtc tgt cgg caa gga ggg acg cat gcc act ctc cga tca tga gca gcg gat
 tyr val thr val cys arg gln gly gly thr his ala thr leu arg ser OPA ala ala asp
 1141/381
 gct tga cca gat cga gag cgc tct cta cgc cga aga tcc caa gtt cgc atc gag tgt ccg
 ala OPA pro asp arg glu arg ser leu arg arg arg ser gln val arg ile glu cys pro
 1201/401
 tgg cgg ggg ctt ccg cgc acc gac cgc gcg gcg gcg cct gca ggg cgc ggc gtt gtt cat
 trp arg gly leu pro arg thr asp arg ala ala ala pro ala gly arg gly val val his
 1261/421
 cat cgg tct ggg gat gtt ggt ttc cgg cgt ggc gtt caa aga gac cat gat cgg aag ttt
 his arg ser gly asp val gly phe arg arg gly val gln arg asp his asp arg lys phe
 1321/441
 ccc gat act cag cgt ttt cgg ttt tgt cgt gat gtt cgg tgg tgt ggt gta tgc cat cac
 pro asp thr gln arg phe arg phe cys arg asp val arg trp cys gly val cys his his
 1381/461
 cgg tcc tcg gtt gtc cgg cag gat gga tcg tgg cgg atc ggc tgc tgg ggc ttc gcg cca
 arg ser ser val val arg gln asp gly ser trp arg ile gly cys trp gly phe ala pro
 1441/481
 1471/491
 gcg tcg tac caa ggg ggc cgg ggg ctc att cac cag ccg tat gga aga tc
 ala ser tyr gln gly gly arg gly leu ile his gln pro tyr gly arg

SEQ ID N° 452C (suite 1)

FIGURE 452C (suite 1)

ORF de seq 452A directement en fusion avec phoA
 cag tct gtc ggc aag gag gga cgc atg cca ctc tcc gat cat gag cag cgg
 gln ser val gly lys glu gly arg met pro leu ser asp his glu gln arg
 1141/381
 1171/391
 atg ctt gac cag atc gag agc gct ctc tac gcc gaa gat ccc aag ttc gca tcg agt gtc
 met leu asp gln ile glu ser ala leu tyr ala glu asp pro lys phe ala ser ser val
 1201/401
 1231/411
 cgt ggc ggg ggc ttc cgc gca ccg acc gcg cgg cgg cgc ctg cag ggc gcg gcg ttg ttc
 arg gly gly gly phe arg ala pro thr ala arg arg arg leu gln gly ala ala leu phe
 1261/421
 1291/431
 atc atc ggt ctg ggg atg ttg gtt tcc ggc gtg gcg ttc aaa gag acc atg atc gga agt
 ile ile gly leu gly met leu val ser gly val ala phe lys glu thr met ile gly ser
 1321/441
 1351/451
 ttc ccg ata ctc agc gtt ttc ggt ttt gtc gtg atg ttc ggt ggt gtg gtg tat gcc atc
 phe pro ile leu ser val phe gly phe val val met phe gly gly val val tyr ala ile
 1381/461
 1411/471
 acc ggt cct cgg ttg tcc ggc agg atg gat cgt ggc gga tcg gct gct ggg gct tcg cgc
 thr gly pro arg leu ser gly arg met asp arg gly gly ser ala ala gly ala ser arg
 1441/481
 1471/491
 cag cgt cgt acc aag ggg gcc ggg ggc tca ttc acc agc cgt atg gaa gat c
 gln arg arg thr lys gly ala gly gly ser phe thr ser arg met glu asp

SEQ ID N° 45A

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 45A

153/185

Séquence Rv2169c prédite par Cole et al., 1998 (Nature 393:537-544) et contenant Seq45A

```

1/1                                     31/11
atg cca ctc tcc gat cat gag cag cgg atg ctt gac cag atc gag agc gct ctc tac gcc
Met pro leu ser asp his glu gln arg met leu asp gln ile glu ser ala leu tyr ala
61/21                                     91/31
gaa gat ccc aag ttc gca tcg agt gtc cgt ggc ggg ggc ttc cgc gca ccg acc gcg cgg
glu asp pro lys phe ala ser ser val arg gly gly phe arg ala pro thr ala arg
121/41                                     151/51
cgg cgc ctg cag ggc gcg gcg ttg ttc atc atc ggt ctg ggg atg ttg gtt tcc ggc gtg
arg arg leu gln gly ala ala leu phe ile ile gly leu gly met leu val ser gly val
181/61                                     211/71
gcg ttc aaa gag acc atg atc gga agt ttc ccg ata ctc agc gtt ttc ggt ttt gtc gtg
ala phe lys glu thr met ile gly ser phe pro ile leu ser val phe gly phe val val
241/81                                     271/91
atg ttc ggt ggt gtg gtg tat gcc atc acc ggt cct cgg ttg tcc ggc agg atg gat cgt
met phe gly gly val val tyr ala ile thr gly pro arg leu ser gly arg met asp arg
301/101                                    331/111
ggc gga tcg gct gct ggg gct tcg cgc cag cgt cgt acc aag ggg gcc ggg ggc tca ttc
gly gly ser ala ala gly ala ser arg gln arg arg thr lys gly ala gly gly ser phe
361/121                                    391/131
acc agc cgt atg gaa gat cgg ttc cgg cgc cgc ttc gac gag taa
thr ser arg met glu asp arg phe arg arg arg phe asp glu OCH

```

SEQ ID N° 45D

FIGURE 45D

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv2169c

```

1/1                                     31/11
tga cag tct gtc ggc aag gag gga cgc atg cca ctc tcc gat cat gag cag cgg atg ctt
OPA gln ser val gly lys glu gly arg met pro leu ser asp his glu gln arg met leu
61/21                                     91/31
gac cag atc gag agc gct ctc tac gcc gaa gat ccc aag ttc gca tcg agt gtc cgt ggc
asp gln ile glu ser ala leu tyr ala glu asp pro lys phe ala ser ser val arg gly
121/41                                     151/51
ggg ggc ttc cgc gca ccg acc gcg cgg cgg cgc ctg cag ggc gcg gcg ttg ttc atc atc
gly gly phe arg ala pro thr ala arg arg arg leu gln gly ala ala leu phe ile ile
181/61                                     211/71
ggt ctg ggg atg ttg gtt tcc ggc gtg gcg ttc aaa gag acc atg atc gga agt ttc ccg
gly leu gly met leu val ser gly val ala phe lys glu thr met ile gly ser phe pro
241/81                                     271/91
ata ctc agc gtt ttc ggt ttt gtc gtg atg ttc ggt ggt gtg gtg tat gcc atc acc ggt
ile leu ser val phe gly phe val val met phe gly gly val val tyr ala ile thr gly
301/101                                    331/111
cct cgg ttg tcc ggc agg atg gat cgt ggc gga tcg gct gct ggg gct tcg cgc cag cgt
pro arg leu ser gly arg met asp arg gly gly ser ala ala gly ala ser arg gln arg
361/121                                    391/131
cgt acc aag ggg gcc ggg ggc tca ttc acc agc cgt atg gaa gat cgg ttc cgg cgc cgc
arg thr lys gly ala gly gly ser phe thr ser arg met glu asp arg phe arg arg arg
421/141
ttc gac gag taa
phe asp glu OCH

```

SEQ ID N° 45F

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 45F

154/185

```

1/1                                31/11
cag ccg cgc cgc atc gac cag ggc ctc acg ccc ggt cac ttc tcc gcg ttc ctc aac aat
gln pro arg arg ile asp gln gly leu thr pro gly his phe ser ala phe leu asn asn
61/21                                91/31
tcc ggt gaa cat cgc acc agg tta ggc agc aat ccc gcg gac ccg cac ccc act cgc cga
ser gly glu his arg thr arg leu gly ser asn pro ala asp pro his pro thr arg arg
121/41                                151/51
ccg gcc aac tca cag aca ccc tct acg atg cag ggt atg cgg acc ccc aga cgc cac tgc
pro ala asn ser gln thr pro ser thr met gln gly met arg thr pro arg arg his cys
181/61                                211/71
cgt cgc atc gcc gtc ctc gcc gcc gtt agc atc gcc gcc act gtc gtt gcc gcc tgc tcg
arg arg ile ala val leu ala ala val ser ile ala ala thr val val ala gly cys ser
241/81                                271/91
tcg gcc tcg aag cca agc gcc gga cca ctt ccg gac gcg aag ccg ctg gtc gag gag gcc
ser gly ser lys pro ser gly gly pro leu pro asp ala lys pro leu val glu glu ala
301/101                                331/111
acc gcg cag acc aag gct ctc aag agc gcg cac atg gtg ctg acg gtc aac gcc aag atc
thr ala gln thr lys ala leu lys ser ala his met val leu thr val asn gly lys ile

```

SEQ ID N° 46A

FIGURE 46A

```

1/1                                31/11
agc cgc gcc gca tcg acc agg gcc tca cgc ccg gtc act tct ccg cgt tcc tca aca att
ser arg ala ala ser thr arg ala ser arg pro val thr ser pro arg ser ser thr ile
61/21                                91/31
ccg gtg aac atc gca cca ggt tag gca gca atc ccg ccg acc cgc acc cca ctc gcc gac
pro val asn ile ala pro gly AMB ala ala ile pro arg thr arg thr pro leu ala asp
121/41                                151/51
cgg cca act cac aga cac cct cta cga tgc agg gta tgc gga ccc cca gac gcc act gcc
arg pro thr his arg his pro leu arg cys arg val cys gly pro pro asp ala thr ala
181/61                                211/71
gtc gca tcg ccg tcc tcg ccg ccg tta gca tcg ccg cca ctg tcg ttg ccg gct gct cgt
val ala ser pro ser ser pro pro leu ala ser pro pro leu ser leu pro ala ala arg
241/81                                271/91
cgg gct cga agc caa gcg gcg gac cac ttc ccg acg cga agc cgc tgg tcg agg agg cca
arg ala arg ser gln ala ala asp his phe arg thr arg ser arg trp ser arg arg pro
301/101                                331/111
ccg cgc aga cca agg ctc tca aga gcg cgc aca tgg tgc tga ccg tca acg gca aga tc
pro arg arg pro arg leu ser arg ala arg thr trp cys OPA arg ser thr ala arg

```

SEQ ID N° 46B

FIGURE 46B

155/185

1/1	31/11
gcc gcg ccg cat cga cca ggg cct cac gcc	cgg tca ctt ctc cgc gtt cct caa caa ttc
ala ala pro his arg pro gly pro his ala	arg ser leu leu arg val pro gln gln phe
61/21	91/31
cgg tga aca tcg cac cag gtt agg cag caa	tcc cgc gga ccc gca ccc cac tcg ccg acc
arg opa thr ser his gln val arg gln gln	ser arg gly pro ala pro his ser pro thr
121/41	151/51
ggc caa ctc aca gac acc ctc tac gat gca	ggg tat gcg gac ccc cag acg cca ctg ccg
gly gln leu thr asp thr leu tyr asp ala	gly tyr ala asp pro gln thr pro leu pro
181/61	211/71
tcg cat cgc cgt cct cgc cgc cgt tag cat	cgc cgc cac tgt cgt tgc cgg ctg ctc gtc
ser his arg arg pro arg arg arg AMB his	arg arg his cys arg cys arg leu leu val
241/81	271/91
ggg ctc gaa gcc aag cgg cgg acc act tcc	gga cgc gaa gcc gct ggt cga gga ggc cac
gly leu glu ala lys arg arg thr thr ser	gly arg glu ala ala gly arg gly gly his
301/101	331/111
cgc gca gac caa ggc tct caa gag cgc gca	cat ggt gct gac ggt caa cgg caa gat c
arg ala asp gln gly ser gln glu arg ala	his gly ala asp gly gln arg gln asp

SEQ ID N° 46C

FIGURE 46C

156/185

Séquence codante Rv1411c prédite par Cole et al., 1998 (Nature 393: 537-544) et contenant seq46A:

```

1/1                               31/11
atg cgg acc ccc aga cgc cac tgc cgt cgc atc gcc gtc ctc gcc gcc gtt agc atc gcc
Met arg thr pro arg arg his cys arg arg ile ala val leu ala ala val ser ile ala
61/21                               91/31
gcc act gtc gtt gcc ggc tgc tcg tcg ggc tcg aag cca agc ggc gga cca ctt ccg gac
ala thr val val ala gly cys ser ser gly ser lys pro ser gly gly pro leu pro asp
121/41                               151/51
gcg aag ccg ctg gtc gag gag gcc acc gcg cag acc aag gct ctc aag agc gcg cac atg
ala lys pro leu val glu glu ala thr ala gln thr lys ala leu lys ser ala his met
181/61                               211/71
gtg ctg acg gtc aac ggc aag atc ccg gga ctg tct ctg aag acg ctg agc ggc gat ctc
val leu thr val asn gly lys ile pro gly leu ser leu lys thr leu ser gly asp leu
241/81                               271/91
acc acc aac ccc acc gcc gcg acg gga aac gtc aag ctc acg ctg ggt ggg tct gat atc
thr thr asn pro thr ala ala thr gly asn val lys leu thr leu gly gly ser asp ile
301/101                               331/111
gat gcc gac ttc gtg gtg ttc gac ggg atc ctg tac gcc acc ctg acg ccc aac cag tgg
asp ala asp phe val val phe asp gly ile leu tyr ala thr leu thr pro asn gln trp
361/121                               391/131
agc gat ttc ggt ccc gcc gcc gac atc tac gac ccc gcc cag gtg ctg aat ccg gat acc
ser asp phe gly pro ala ala asp ile tyr asp pro ala gln val leu asn pro asp thr
421/141                               451/151
ggc ctg gcc aac gtg ctg gcg aat ttc gcc gac gca aaa gcc gaa ggg cgg gat acc atc
gly leu ala asn val leu ala asn phe ala asp ala lys ala glu gly arg asp thr ile
481/161                               511/171
aac ggc cag aac acc atc cgc atc agc ggg aag gta tcg gca cag gcg gtg aac cag ata
asn gly gln asn thr ile arg ile ser gly lys val ser ala gln ala val asn gln ile
541/181                               571/191
gcg ccg ccg ttc aac gcg acg cag ccg gtg ccg gcg acc gtc tgg att cag gag acc ggc
ala pro pro phe asn ala thr gln pro val pro ala thr val trp ile gln glu thr gly
601/201                               631/211
gat cat caa ctg gca cag gcc cag ttg gac cgc ggc tcg ggc aat tcc gtc cag atg acc
asp his gln leu ala gln ala gln leu asp arg gly ser gly asn ser val gln met thr
661/221                               691/231
ttg tcg aaa tgg ggc gag aag gtc cag gtc acg aag ccc ccg gtg agc tga
leu ser lys trp gly glu lys val gln val thr lys pro pro val ser OPA

```

SEQ ID N° 46D

FIGURE 46D

157/185

ORF d'après par Cole et al., 1998 (Nature 393: 537-544):
et contenant la séquence codante Rv1411c:

```

1/1                                31/11
tag ctc acc cag gtt gga ccg gtt cag tgt ctc ggc cat cac gtc ggc ggt gaa ttg gcc
AMB leu thr gln val gly pro val gln cys leu gly his his val gly gly glu leu ala
61/21                                91/31
gtc ggg caa tac atc gac gac cgt cag aca cac gcc gtt gac agc gat cga gtc gcc gtg
val gly gln tyr ile asp asp arg gln thr his ala val asp ser asp arg val ala val
121/41                                151/51
gcc ggc gtc ggc ggt aac cat cgg acc gcg gat ggt cag ccg cgc cgc atc gac cag ggc
ala gly val gly gly asn his arg thr ala asp gly gln pro arg arg ile asp gln gly
181/61                                211/71
ctc acg ccc ggt cac ttc tcc gcg ttc ctc aac aat tcc ggt gaa cat cgc acc agg tta
leu thr pro gly his phe ser ala phe leu asn asn ser gly glu his arg thr arg leu
241/81                                271/91
ggc agc aat ccc gcg gac ccg cac ccc act cgc cga ccg gcc aac tca cag aca ccc tct
gly ser asn pro ala asp pro his pro thr arg arg pro ala asn ser gln thr pro ser
301/101                                331/111
acg atg cag ggt atg cgg acc ccc aga cgc cac tgc cgt cgc atc gcc gtc ctc gcc gcc
thr met gln gly met arg thr pro arg arg his cys arg arg ile ala val leu ala ala
361/121                                391/131
gtt agc atc gcc gcc act gtc gtt gcc ggc tgc tgc tgc ggc tgc aag cca agc ggc gga
val ser ile ala ala thr val val ala gly cys ser ser gly ser lys pro ser gly gly
421/141                                451/151
cca ctt ccg gac gcg aag ccg ctg gtc gag gag gcc acc gcg cag acc aag gct ctc aag
pro leu pro asp ala lys pro leu val glu glu ala thr ala gln thr lys ala leu lys
481/161                                511/171
agc gcg cac atg gtg ctg acg gtc aac ggc aag atc ccg gga ctg tct ctg aag acg ctg
ser ala his met val leu thr val asn gly lys ile pro gly leu ser leu lys thr leu
541/181                                571/191
agc ggc gat ctc acc acc aac ccc acc gcc gcg acg gga aac gtc aag ctc acg ctg ggt
ser gly asp leu thr thr asn pro thr ala ala thr gly asn val lys leu thr leu gly
601/201                                631/211
ggg tct gat atc gat gcc gac ttc gtg gtg ttc gac ggg atc ctg tac gcc acc ctg acg
gly ser asp ile asp ala asp phe val val phe asp gly ile leu tyr ala thr leu thr
661/221                                691/231
ccc aac cag tgg agc gat ttc ggt ccc gcc gcc gac atc tac gac ccc gcc cag gtg ctg
pro asn gln trp ser asp phe gly pro ala ala asp ile tyr asp pro ala gln val leu
721/241                                751/251
aat ccg gat acc ggc ctg gcc aac gtg ctg gcg aat ttc gcc gac gca aaa gcc gaa ggg
asn pro asp thr gly leu ala asn val leu ala asn phe ala asp ala lys ala glu gly
781/261                                811/271
cgg gat acc atc aac ggc cag aac acc atc cgc atc agc ggg aag gta tgc gca cag gcg
arg asp thr ile asn gly gln asn thr ile arg ile ser gly lys val ser ala gln ala
841/281                                871/291
gtg aac cag ata gcg ccg ccg ttc aac gcg acg cag ccg gtg ccg gcg acc gtc tgg att
val asn gln ile ala pro pro phe asn ala thr gln pro val pro ala thr val trp ile
901/301                                931/311
cag gag acc ggc gat cat caa ctg gca cag gcc cag ttg gac cgc ggc tgc ggc aat tcc
gln glu thr gly asp his gln leu ala gln ala gln leu asp arg gly ser gly asn ser
961/321                                991/331
gtc cag atg acc ttg tgc aaa tgg ggc gag aag gtc cag gtc acg aag ccc ccg gtg agc
val gln met thr leu ser lys trp gly glu lys val gln val thr lys pro pro val ser
1021/341
tga
OPA

```

SEQ ID N° 46F

FEUILLE DE REMPLACEMENT (REGLE 26)

158/185

```

1/1                               31/11
gag ctg gtc aac ggc gcc ggc atc gac gac gcc gcc gtc gtg acc tgc cgg ccg gac agc
glu leu val asn gly ala gly ile asp asp ala ala val val thr cys arg pro asp ser
61/21                               91/31
ctg gcc gat gcc cag cag atg gtc gag gcg gca ctg ggc cga tat ggc cgt ttg gac gga
leu ala asp ala gln gln met val glu ala ala leu gly arg tyr gly arg leu asp gly
121/41                               151/51
gtg ttg gtg gcc tcg ggc agc aac cat gtg gcg ccc att acc gag atg gcc gtc gag gac
val leu val ala ser gly ser asn his val ala pro ile thr glu met ala val glu asp
181/61                               211/71
ttc gac gct gtg atg gac gcg aac gtg gcg ggt gcc tgg ctg gtg tgt cgg gcg gcc gga
phe asp ala val met asp ala asn val arg gly ala trp leu val cys arg ala ala gly
241/81                               271/91
cgg gtg ctg ctc gag cag ggt cag ggc ggc agc gtg gtg ctg gtg tcg tcc gtt cgc ggc
arg val leu leu glu gln gly gln gly gly ser val val leu val ser ser val arg gly
301/101                               331/111
ggg ttg ggc aat gcc gcc ggt tac agc gcg tac tgc ccg tcg aag gcg ggc acc gat c
gly leu gly asn ala ala gly tyr ser ala tyr cys pro ser lys ala gly thr asp

```

SEQ ID N° 47A

FIGURE 47A

```

1/1                               31/11
agc tgg tca acg gcg ccg gca tcg acg acg ccg ccg tcg tga cct gcc ggc cgg aca gcc
ser trp ser thr ala pro ala ser thr thr pro pro ser OPA pro ala gly arg thr ala
61/21                               91/31
tgg ccg atg ccc agc aga tgg tcg agg ccg cac tgg gcc gat atg gcc gtt tgg acg gag
trp pro met pro ser arg trp ser arg arg his trp ala asp met ala val trp thr glu
121/41                               151/51
tgt tgg tgg cct ccg gca gca acc atg tgg cgc cca tta ccg aga tgg ccg tcg agg act
cys trp trp pro arg ala ala thr met trp arg pro leu pro arg trp pro ser arg thr
181/61                               211/71
tcg acg ctg tga tgg acg cga acg tgc ggc gtg cct ggc tgg tgt gtc ggc ccg ccg gac
ser thr leu OPA trp thr arg thr cys gly val pro gly trp cys val gly arg pro asp
241/81                               271/91
ggg tgc tgc tcg agc agg gtc agg gcg gca gcg tgg tgc tgg tgt cgt ccg ttc gcg gcg
gly cys cys ser ser arg val arg ala ala ala trp cys trp cys arg pro phe ala ala
301/101                               331/111
ggt tgg gca atg ccg ccg gtt aca gcg cgt act gcc cgt cga agg ccg gca ccg atc
gly trp ala met pro pro val thr ala arg thr ala arg arg arg ala pro ile

```

SEQ ID N° 47B

FIGURE 47B

159/185

```

1/1                                31/11
gct ggt caa cgg cgc cgg cat cga cga cgc cgc cgt cgt gac ctg ccg gcc gga cag cct
ala gly gln arg arg arg his arg arg arg arg arg arg asp leu pro ala gly gln pro
61/21                                91/31
ggc cga tgc cca gca gat ggt cga ggc ggc act ggg ccg ata tgg ccg ttt gga cgg agt
gly arg cys pro ala asp gly arg gly gly thr gly pro ile trp pro phe gly arg ser
121/41                                151/51
gtt ggt ggc ctc ggg cag caa cca tgt ggc gcc cat tac cga gat ggc cgt cga gga ctt
val gly gly leu gly gln gln pro cys gly ala his tyr arg asp gly arg arg gly leu
181/61                                211/71
cga cgc tgt gat gga cgc gaa cgt gcg ggg tgc ctg gct ggt gtg tcg ggc ggc cgg acg
arg arg cys asp gly arg glu arg ala gly cys leu ala gly val ser gly gly arg thr
241/81                                271/91
ggt gct gct cga gca ggg tca ggg cgg cag cgt ggt gct ggt gtc gtc cgt tcg cgg cgg
gly ala ala arg ala gly ser gly arg gln arg gly ala gly val val arg ser arg arg
301/101                                331/111
gtt ggg caa tgc cgc cgg tta cag cgc gta ctg ccc gtc gaa ggc ggg cac cga tc
val gly gln cys arg arg leu gln arg val leu pro val glu gly gly his arg

```

SEQ ID N° 47C

FIGURE 47C

160/185

Séquence codante Rv1714 prédite par Cole et al., 1998 (Nature 393: 537-544) et contenant seq 47A:

```

1/1                               31/11
gtg gag gaa atg gcg ctg gct cag cag gtg ccg aac ctg ggt ctg gcg cgc ttc agc gtg
val glu glu met ala leu ala gln gln val pro asn leu gly leu ala arg phe ser val
61/21                               91/31
cag gac aag tcg atc ctg atc acc ggc gcg acc ggt tcg ttg ggc cga gtt gcc gcc cgg
gln asp lys ser ile leu ile thr gly ala thr gly ser leu gly arg val ala ala arg
121/41                               151/51
gcg ctg gcc gac gcg gga gcg cgg ctg aca ctg gcc ggc ggc aac tcg gcc ggt ctg gcc
ala leu ala asp ala gly ala arg leu thr leu ala gly gly asn ser ala gly leu ala
181/61                               211/71
gag ctg gtc aac ggc gcc ggc atc gac gac gcc gcc gtc gtg acc tgc cgg ccg gac agc
glu leu val asn gly ala gly ile asp asp ala ala val val thr cys arg pro asp ser
241/81                               271/91
ctg gcc gat gcc cag cag atg gtc gag gcg gca ctg ggc cga tat ggc cgt ttg gac gga
leu ala asp ala gln gln met val glu ala ala leu gly arg tyr gly arg leu asp gly
301/101                               331/111
gtg ttg gtg gcc tcg ggc agc aac cat gtg gcg ccc att acc gag atg gcc gtc gag gac
val leu val ala ser gly ser asn his val ala pro ile thr glu met ala val glu asp
361/121                               391/131
ttc gac gct gtg atg gac gcg aac gtg cgg ggt gcc tgg ctg gtg tgt cgg gcg gcc gga
phe asp ala val met asp ala asn val arg gly ala trp leu val cys arg ala ala gly
421/141                               451/151
cgg gtg ctg ctc gag cag ggt cag ggc ggc agc gtg gtg ctg gtg tcg tcc gtt cgc ggc
arg val leu leu glu gln gly gln gly gly ser val val leu val ser ser val arg gly
481/161                               511/171
ggg ttg ggc aat gcc gcc ggt tac agc gcg tac tgc ccg tcg aag gcg ggc acc gat ctg
gly leu gly asn ala ala gly tyr ser ala tyr cys pro ser lys ala gly thr asp leu
541/181                               571/191
ttg gcc aag aca ttg gcg gcc gaa tgg ggc ggt cac ggc att cgg gtg aac gcg ctg gcg
leu ala lys thr leu ala ala glu trp gly gly his gly ile arg val asn ala leu ala
601/201                               631/211
ccg acg gtg ttt cgg tcc gcg gtg acc gag tgg atg ttc acc gac gat ccg aag ggc cgg
pro thr val phe arg ser ala val thr glu trp met phe thr asp asp pro lys gly arg
661/221                               691/231
gcc acc cgg gag gcg atg ctc gcc cgg atc ccg ttg cgc cgc ttc gcc gaa ccg gaa gac
ala thr arg glu ala met leu ala arg ile pro leu arg arg phe ala glu pro glu asp
721/241                               751/251
ttc gtc ggc gcc ctg atc tat ctg ctc agc gac gcc tcg agc ttc tac acc ggc cag gtg
phe val gly ala leu ile tyr leu leu ser asp ala ser ser phe tyr thr gly gln val
781/261                               811/271
atg tat ctg gac ggc ggg tac acc gca tgc tga
met tyr leu asp gly gly tyr thr ala cys OPA

```

SEQ ID N° 47D

FIGURE 47D

161/185

ORF d'après Cole et al., 1998 (Nature 393: 537-544) et contenant la séquence codante Rv1714:

24/1	54/11
tag gtg gag gaa atg gcg ctg gct cag cag	gtg ccg aac ctg ggt ctg gcg cgc ttc agc
AMB val glu glu met ala leu ala gln gln	val pro asn leu gly leu ala arg phe ser
84/21	114/31
gtg cag gac aag tcg atc ctg atc acc ggc	gcg acc ggt tcg ttg ggc cga gtt gcc gcc
val gln asp lys ser ile leu ile thr gly	ala thr gly ser leu gly arg val ala ala
144/41	174/51
cgg gcg ctg gcc gac gcg gga gcg cgg ctg	aca ctg gcc ggc ggc aac tcg gcc ggt ctg
arg ala leu ala asp ala gly ala arg leu	thr leu ala gly gly asn ser ala gly leu
204/61	234/71
gcc gag ctg gtc aac ggc gcc ggc atc gac	gac gcc gcc gtc gtg acc tgc cgg ccg gac
ala glu leu val asn gly ala gly ile asp	asp ala ala val val thr cys arg pro asp
264/81	294/91
agc ctg gcc gat gcc cag cag atg gtc gag	gcg gca ctg ggc cga tat ggc cgt ttg gac
ser leu ala asp ala gln gln met val glu	ala ala leu gly arg tyr gly arg leu asp
324/101	354/111
gga gtg ttg gtg gcc tcg ggc agc aac cat	gtg gcg ccc att acc gag atg gcc gtc gag
gly val leu val ala ser gly ser asn his	val ala pro ile thr glu met ala val glu
384/121	414/131
gac ttc gac gct gtg atg gac gcg aac gtg	cgg ggt gcc tgg ctg gtg tgt cgg gcg gcc
asp phe asp ala val met asp ala asn val	arg gly ala trp leu val cys arg ala ala
444/141	474/151
gga cgg gtg ctg ctc gag cag ggt cag ggc	ggc agc gtg gtg ctg gtg tcg tcc gtt cgc
gly arg val leu leu glu gln gly gln gly	gly ser val val leu val ser ser val arg
504/161	534/171
ggc ggg ttg ggc aat gcc gcc ggt tac agc	gcg tac tgc ccg tcg aag gcg ggc acc gat
gly gly leu gly asn ala ala gly tyr ser	ala tyr cys pro ser lys ala gly thr asp
564/181	594/191
ctg ttg gcc aag aca ttg gcg gcc gaa tgg	ggc ggt cac ggc att cgg gtg aac gcg ctg
leu leu ala lys thr leu ala ala glu trp	gly gly his gly ile arg val asn ala leu
624/201	654/211
gcg ccg acg gtg ttt cgg tcc gcg gtg acc	gag tgg atg ttc acc gac gat ccg aag ggc
ala pro thr val phe arg ser ala val thr	glu trp met phe thr asp asp pro lys gly
684/221	714/231
cgg gcc acc cgg gag gcg atg ctc gcc cgg	atc ccg ttg cgc cgc ttc gcc gaa ccg gaa
arg ala thr arg glu ala met leu ala arg	ile pro leu arg arg phe ala glu pro glu
744/241	774/251
gac ttc gtc ggc gcc ctg atc tat ctg ctc	agc gac gcc tcg agc ttc tac acc ggc cag
asp phe val gly ala leu ile tyr leu leu	ser asp ala ser ser phe tyr thr gly gln
804/261	834/271
gtg atg tat ctg gac ggc ggg tac acc gca	tgc tga
val met tyr leu asp gly gly tyr thr ala	cys OPA

SEQ ID N° 47F

FIGURE 47F

162/185

```

1/1                               31/11
agg ctc atg agc aag acg gtt ctc atc ctt ggc gcg ggt gtc ggc ggc ctg acc acc gcc
arg leu met ser lys thr val leu ile leu gly ala gly val gly gly leu thr thr ala
61/21                             91/31
gac acc ctc cgt caa ctg cta cca cct gag gat c
asp thr leu arg gln leu leu pro pro glu asp

```

SEQ ID N° 48A

FIGURE 48A

```

1/1                               31/11
ggc tca tga gca aga cgg ttc tca tcc ttg gcg cgg gtg tcg gcg gcc tga cca ccg ccg
gly ser OPA ala arg arg phe ser ser leu ala arg val ser ala ala OPA pro pro pro
61/21                             91/31
aca ccc tcc gtc aac tgc tac cac ctg agg atc
thr pro ser val asn cys tyr his leu arg ile

```

SEQ ID N° 48B

FIGURE 48B

```

1/1                               31/11
gct cat gag caa gac ggt tct cat cct tgg cgc ggg tgt cgg cgg cct gac cac cgc cga
ala his glu gln asp gly ser his pro trp arg gly cys arg arg pro asp his arg arg
61/21
cac cct ccg tca act gct acc acc tga gga tc
his pro pro ser thr ala thr thr OPA gly

```

SEQ ID N° 48C

FIGURE 48C

163/185

Séquence codante Rv0331 prédite par Cole et al., 1998 (Nature 393: 537-544) et contenant seq48A:

```

1/1                               31/11
atg agc aag acg gtt ctc atc ctt ggc gcg ggt gtc ggc ggc ctg acc acc gcc gac acc
Met ser lys thr val leu ile leu gly ala gly val gly gly leu thr thr ala asp thr
61/21                               91/31
ctc cgt caa ctg cta cca cct gag gat cga atc ata ttg gtg gac agg agc ttt gac ggg
leu arg gln leu leu pro pro glu asp arg ile ile leu val asp arg ser phe asp gly
121/41                               151/51
acg ctg ggc ttg tcg ttg cta tgg gtg ttg cgg ggc tgg cgg cgg cct gac gac gtc cgc
thr leu gly leu ser leu leu trp val leu arg gly trp arg arg pro asp asp val arg
181/61                               211/71
gtc cgc ccc acc gcg gcg tcg ctg ccc ggt gtg gaa atg gtt act gca acc gtc gcc cac
val arg pro thr ala ala ser leu pro gly val glu met val thr ala thr val ala his
241/81                               271/91
att gac atc gcg gcc cag gta gtg cac acc gac aac agc gtc atc ggc tat gac gcg ttg
ile asp ile ala ala gln val val his thr asp asn ser val ile gly tyr asp ala leu
301/101                               331/111
gtg atc gca tta ggt gcg gcg ctg aac acc gac gcc gtt ccc gga ctg tcg gac gcg ctc
val ile ala leu gly ala ala leu asn thr asp ala val pro gly leu ser asp ala leu
361/121                               391/131
gac gcc gac gtc gcg ggc cag ttc tac acc ctg gac ggc gcg gct gag ctg cgt gcg aag
asp ala asp val ala gly gln phe tyr thr leu asp gly ala ala glu leu arg ala lys
421/141                               451/151
gtc gag gcg ctc gag cat ggc cgg atc gct gtg gct atc gcc ggg gtg ccg ttc aaa tgc
val glu ala leu glu his gly arg ile ala val ala ile ala gly val pro phe lys cys
481/161                               511/171
cca gcc gca ccg ttc gaa gcg gcg ttt ctg atc gcc gcc caa ctc ggt gac cgc tac gcc
pro ala ala pro phe glu ala ala phe leu ile ala ala gln leu gly asp arg tyr ala
541/181                               571/191
acc gga acc gta cag atc gac acg ttc acg cct gac ccg ctg ccg atg ccc gtt gca ggt
thr gly thr val gln ile asp thr phe thr pro asp pro leu pro met pro val ala gly
601/201                               631/211
ccc gag gtc ggc gag gct ttg gtc tcg atg ctc aag gat cac ggt gtc ggc ttc cat cct
pro glu val gly glu ala leu val ser met leu lys asp his gly val gly phe his pro
661/221                               691/231
cgc aag gcc cta gct cgc gtc gat gag gcc gca agg acg atg cac ttc ggt gac ggc acg
arg lys ala leu ala arg val asp glu ala ala arg thr met his phe gly asp gly thr
721/241                               751/251
tcc gaa ccg ttc gat ctg ctt gcc gtg gtc ccc ccg cac gtg ccc tcc gcc gcg gcg cgg
ser glu pro phe asp leu leu ala val val pro pro his val pro ser ala ala ala arg
781/261                               811/271
tca gcg ggt ctc agc gaa tcc ggg tgg ata ccc gtg gac ccg cgc acc ctg tcc act agc
ser ala gly leu ser glu ser gly trp ile pro val asp pro arg thr leu ser thr ser
841/281                               871/291
gcc gac aac gtg tgg gcc atc ggc gat gcg acc gtg ctg acg ctg ccg aat ggc aaa ccg
ala asp asn val trp ala ile gly asp ala thr val leu thr leu pro asn gly lys pro
901/301                               931/311
ctg ccc aag gct gcc gtg ttc gcc gaa gcc cag gcc gca gtt gtc gcc cac ggc gtc gcc
leu pro lys ala ala val phe ala glu ala gln ala ala val val ala his gly val ala
961/321                               991/331
cgc cat ctc ggt tac gac gta gct gag cgc cac ttc acc ggc acg ggc gcc tgc tac gtc
arg his leu gly tyr asp val ala glu arg his phe thr gly thr gly ala cys tyr val
1021/341                               1051/351
gag acc ggt gat cac cag gca gcc aag ggc gac ggc gat ttc ttc gct ccg tcg gcg ccc
glu thr gly asp his gln ala ala lys gly asp gly asp phe phe ala pro ser ala pro
1081/361                               1111/371
tcg gtg acg ctg tac ccg ccg tcg cgg gag ttt cac gag gag aag gtc gca caa gaa ctg
ser val thr leu tyr pro pro ser arg glu phe his glu glu lys val ala gln glu leu
1141/381
gcc tgg ctg acc cgc tgg aag acg tga
ala trp leu thr arg trp lys thr OPA.

```

SEQ ID N° 48D

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 48D

164/185

ORF d'après Cole et al., 1998 (Nature 393: 537-544) et contenant la séquence codante Rv0331:

1/1	31/11
tga aca ccc gcg ccg acg cgg cga caa tcg	cgg aaa acc ggt ccg cgg gaa tgc tgc ggg
OPA thr pro ala pro thr arg arg gln ser	arg lys thr gly pro arg glu cys cys gly
61/21	91/31
cca tgg gcc gat aat agt ttg act gac tcg	gtc agt cac ccc aag acc ttg cgc aag act
pro trp ala asp asn ser leu thr asp ser	val ser his pro lys thr leu arg lys thr
121/41	151/51
gcg gcg gaa tct aat att cca aag ata tat	gga act cga tgc gaa gga atc agg ctc atg
ala ala glu ser asn ile pro lys ile tyr	gly thr arg cys glu gly ile arg leu met
181/61	211/71
agc aag acg gtt ctc atc ctt ggc gcg ggt	gtc ggc ggc ctg acc acc gcc gac acc ctc
ser lys thr val leu ile leu gly ala gly	val gly gly leu thr thr ala asp thr leu
241/81	271/91
cgt caa ctg cta cca cct gag gat cga atc	ata ttg gtg gac agg agc ttt gac ggg acg
arg gln leu leu pro pro glu asp arg ile	ile leu val asp arg ser phe asp gly thr
301/101	331/111
ctg ggc ttg tcg ttg cta tgg gtg ttg cgg	ggc tgg cgg cgg cct gac gac gtc cgc gtc
leu gly leu ser leu leu trp val leu arg	gly trp arg arg pro asp asp val arg val
361/121	391/131
cgc ccc acc gcg gcg tcg ctg ccc ggt gtg	gaa atg gtt act gca acc gtc gcc cac att
arg pro thr ala ala ser leu pro gly val	glu met val thr ala thr val ala his ile
421/141	451/151
gac atc gcg gcc cag gta gtg cac acc gac	aac agc gtc atc ggc tat gac gcg ttg gtg
asp ile ala ala gln val val his thr asp	asn ser val ile gly tyr asp ala leu val
481/161	511/171
atc gca tta ggt gcg gcg ctg aac acc gac	gcc gtt ccc gga ctg tcg gac gcg ctc gac
ile ala leu gly ala ala leu asn thr asp	ala val pro gly leu ser asp ala leu asp
541/181	571/191
gcc gac gtc gcg ggc cag ttc tac acc ctg	gac ggc gcg gct gag ctg cgt gcg aag gtc
ala asp val ala gly gln phe tyr thr leu	asp gly ala ala glu leu arg ala lys val
601/201	631/211
gag gcg ctc gag cat ggc cgg atc gct gtg	gct atc gcc ggg gtg ccg ttc aaa tgc cca
glu ala leu glu his gly arg ile ala val	ala ile ala gly val pro phe lys cys pro
661/221	691/231
gcc gca ccg ttc gaa gcg gcg ttt ctg atc	gcc gcc caa ctc ggt gac cgc tac gcc acc
ala ala pro phe glu ala ala phe leu ile	ala ala gln leu gly asp arg tyr ala thr
721/241	751/251
gga acc gta cag atc gac acg ttc acg cct	gac ccg ctg ccg atg ccc gtt gca ggt ccc
gly thr val gln ile asp thr phe thr pro	asp pro leu pro met pro val ala gly pro
781/261	811/271
gag gtc ggc gag gct ttg gtc tcg atg ctc	aag gat cac ggt gtc ggc ttc cat cct cgc
glu val gly glu ala leu val ser met leu	lys asp his gly val gly phe his pro arg
841/281	871/291
aag gcc cta gct cgc gtc gat gag gcc gca	agg acg atg cac ttc ggt gac ggc acg tcc
lys ala leu ala arg val asp glu ala ala	arg thr met his phe gly asp gly thr ser

SEQ ID N° 48F

FIGURE 48F

FEUILLE DE REMPLACEMENT (REGLE 26)

165/185

901/301	931/311
gaa ccg ttc gat ctg ctt gcc gtg gtc ccc ccg cac gtg ccc tcc gcc gcg gcg cgg tca	
glu pro phe asp leu leu ala val val pro pro his val pro ser ala ala ala arg ser	
961/321	991/331
gcg ggt ctc agc gaa tcc ggg tgg ata ccc gtg gac ccg cgc acc ctg tcc act agc gcc	
ala gly leu ser glu ser gly trp ile pro val asp pro arg thr leu ser thr ser ala	
1021/341	1051/351
gac aac gtg tgg gcc atc ggc gat gcg acc gtg ctg acg ctg ccg aat ggc aaa ccg ctg	
asp asn val trp ala ile gly asp ala thr val leu thr leu pro asn gly lys pro leu	
1081/361	1111/371
ccc aag gct gcc gtg ttc gcc gaa gcc cag gcc gca gtt gtc gcc cac ggc gtc gcc cgc	
pro lys ala ala val phe ala glu ala gln ala ala val val ala his gly val ala arg	
1141/381	1171/391
cat ctc ggt tac gac gta gct gag cgc cac ttc acc ggc acg ggc gcc tgc tac gtc gag	
his leu gly tyr asp val ala glu arg his phe thr gly thr gly ala cys tyr val glu	
1201/401	1231/411
acc ggt gat cac cag gca gcc aag ggc gac ggc gat ttc ttc gct ccg tcg gcg ccc tcg	
thr gly asp his gln ala ala lys gly asp gly asp phe phe ala pro ser ala pro ser	
1261/421	1291/431
gtg acg ctg tac ccg ccg tcg cgg gag ttt cac gag gag aag gtc gca caa gaa ctg gcc	
val thr leu tyr pro pro ser arg glu phe his glu glu lys val ala gln glu leu ala	
1321/441	
tgg ctg acc cgc tgg aag acg tga	
trp leu thr arg trp lys thr OPA	

SEQ ID N° 48F (suite)

FIGURE 48F (suite)

166/185

Fragment amplifié par PCR d'après les similarités de séquences avec une sérine protéase de la famille htrA de E. coli (création du site BamHI à l'extrémité 5' et du site SnaBI à l'extrémité 3') et sous-cloné dans le vecteur pJVED:

```

1/1                                31/11
cca tct aca ccg ctc aac agc cgg gcc aga cgc tgc cgg tcg gtg ctg ccg aga agg cgg
pro ser thr pro leu asn ser arg ala arg arg cys arg ser val leu pro arg arg arg
61/21                                91/31
tga tcc gtg gcg agt tgt tca tgt cgc ggc gca cca ccg ccg acc aac ggg tgc ttg cca
OPA ser val ala ser cys ser cys arg gly ala pro pro pro thr asn gly cys leu pro
121/41                                151/51
tcc gtc tga cca acg gta gtt cgc tgc tga tct cca aaa gtc tca agc cca ccg aag cag
ser val OPA pro thr val val arg cys OPA ser pro lys val ser ser pro pro lys gln
181/61                                211/71
tca tga aca agc tgc gtt ggg tgc tat tga tgc tgg gtg gga tgc ggg tgg cgg tcg ccg
ser OPA thr ser cys val gly cys tyr OPA ser trp val gly ser gly trp arg ser pro
241/81                                271/91
cgg tgg ccg ggg gga tgg tca ccc ggg ccg ggc tga ggc cgg tgg gcc gcc tca ccg aag
arg trp pro gly gly trp ser pro gly pro gly OPA gly arg trp ala ala ser pro lys
301/101                                331/111
cgg ccg agc ggg tgg cgc gaa ccg acg acc tgc ggc cca tcc ccg tct tcg gca gcg acg
arg pro ser gly trp arg glu pro thr thr cys gly pro ser pro ser ser ala ala thr
361/121                                391/131
aat tgg cca ggc tga cag agg cat tca att taa tgc tgc ggg cgc tgg ccg agt cac ggg
asn trp pro gly OPA gln arg his ser ile OCH cys cys gly arg trp pro ser his gly
421/141                                451/151
aac ggc agg caa ggc tgg tta ccg acg ccg gac atg aat tgc gta ccc cgc taa cgt cgc
asn gly arg gln gly trp leu pro thr pro asp met asn cys val pro arg OCH arg arg
481/161                                511/171
tgc gca cca atg tcg aac tct tga tgg cct cga tgg ccc ccg ggg ctc cgc ggc tac cca
cys ala pro met ser asn ser OPA trp pro arg trp pro arg gly leu arg gly tyr pro
541/181                                571/191
agc agg aga tgg tcg acc tgc gtg ccg atg tgc tgg ctc aaa tcg agg aat tgt cca cac
ser arg arg trp ser thr cys val pro met cys trp leu lys ser arg asn cys pro his
601/201                                631/211
tgg tag gcg att tgg tgg acc tgt ccc gag gcg acg ccg gag aag tgg tgc acg agc cgg
trp AMB ala ile trp trp thr cys pro glu ala thr pro glu lys trp cys thr ser arg
661/221                                691/231
tcg aca tgg ctg acg tcg tcg acc gca gcc tgg agc ggg tca ggc ggc ggc gca acg ata
ser thr trp leu thr ser ser thr ala ala trp ser gly ser gly gly gly ala thr ile
721/241                                751/251
tcc ttt tcg acg tcg agg tga ttg ggt ggc agg ttt atg gcg ata ccg ctg gat tgt cgc
ser phe ser thr ser arg OPA leu gly gly arg phe met ala ile pro leu asp cys arg
781/261                                811/271
gga tgg cgc tta acc tga tgg aca acg ccg cga agt gga gcc cgc ccg gcg gcc acg tgg
gly trp arg leu thr OPA trp thr thr pro arg ser gly ala arg arg ala ala thr trp
841/281                                871/291
gtg tca ggc tga gcc agc tcg acg cgt cgc acg ctg agc tgg tgg ttt ccg acc gcg gcc
val ser gly OPA ala ser ser thr arg arg thr leu ser trp trp phe pro thr ala ala

```

SEQ ID N° 49A

FIGURE 49A

FEUILLE DE REMPLACEMENT (REGLE 26)

167/185

901/301	931/311
cgg gca ttc ccg tgc agg agc gcc gtc tgg	tgt ttg aac ggt ttt acc ggt cgg cat cgg
arg ala phe pro cys arg ser ala val trp	cys leu asn gly phe thr gly arg his arg
961/321	991/331
cac ggg cgt tgc cgg gtt cgg gcc tcg ggt	tgg cga tcg tca aac agg tgg tgc tca acc
his gly arg cys arg val arg ala ser gly	trp arg ser ser asn arg trp cys ser thr
1021/341	1051/351
acg gcg gat tgc tgc gca tcg aag aca ccg	acc cag gcg gcc agc ccc ctg gaa cgt cga
thr ala asp cys cys ala ser lys thr pro	thr gln ala ala ser pro leu glu arg arg
1081/361	1111/371
ttt acg tgc tgc tcc ccg gcc gtc gga tgc	cga ttc cgc agc ttc ccg gtg cga cgg ctg
phe thr cys cys ser pro ala val gly cys	arg phe arg ser phe pro val arg arg leu
1141/381	1171/391
gcg ctc gga gca cgg aca tcg aga act ctc	ggg gtt cgg cga acg tta tct cag tgg aat
ala leu gly ala arg thr ser arg thr leu	gly val arg arg thr leu ser gln trp asn
1201/401	1231/411
ctc agt cca cgc gcg caa cct agt tgt gca	gtt act gtt gaa agc cac acc cat gcc agt
leu ser pro arg ala gln pro ser cys ala	val thr val glu ser his thr his ala ser
1261/421	1291/431
cca cgc atg gcc aag ttg gcc cga gta gtg	ggc cta gta cag gaa gag caa cct agc gac
pro arg met ala lys leu ala arg val val	gly leu val gln glu glu gln pro ser asp
1321/441	1351/451
atg acg aat cac cca cgg tat tcg cca ccg	ccg cag cag ccg gga acc cca ggt tat gct
met thr asn his pro arg tyr ser pro pro	pro gln gln pro gly thr pro gly tyr ala
1381/461	1411/471
cag ggg cag cag caa acg tac agc cag cag	ttc gac tgg cgt tac cca ccg tcc ccg ccc
gln gly gln gln gln thr tyr ser gln gln	phe asp trp arg tyr pro pro ser pro pro
1441/481	1471/491
ccg cag cca acc cag tac cgt caa ccc tac	gag gcg ttg ggt ggt acc cgg ccg ggt ctg
pro gln pro thr gln tyr arg gln pro tyr	glu ala leu gly gly thr arg pro gly leu
1501/501	1531/511
ata cct ggc gtg att ccg acc atg acg ccc	cct cct ggg atg gtt cgc caa cgc cct cgt
ile pro gly val ile pro thr met thr pro	pro pro gly met val arg gln arg pro arg
1561/521	1591/531
gca ggc atg ttg gcc atc ggc gcg gtg acg	ata gcg gtg gtg tcc gcc ggc atc ggc ggc
ala gly met leu ala ile gly ala val thr	ile ala val val ser ala gly ile gly gly
1621/541	1651/551
gcg gcc gca tcc ctg gtc ggg ttc aac cgg	gca ccc gcc ggc ccc agc ggc ggc cca gtg
ala ala ala ser leu val gly phe asn arg	ala pro ala gly pro ser gly gly pro val
1681/561	1711/571
gct gcc agc gcg gcg cca agc atc ccc gca	gca aac atg ccg ccg ggg tcg gtc gaa cag
ala ala ser ala ala pro ser ile pro ala	ala asn met pro pro gly ser val glu gln
1741/581	1771/591
gtg gcg gcc aag gtg gtg ccc agt gtc gtc	atg ttg gaa acc gat ctg ggc cgc cag tcg
val ala ala lys val val pro ser val val	met leu glu thr asp leu gly arg gln ser
1801/601	1831/611
gag gag ggc tcc ggc atc att ctg tct gcc	gag ggg ctg atc ttg acc aac aac cac gtg
glu glu gly ser gly ile ile leu ser ala	glu gly leu ile leu thr asn asn his val
1861/621	1891/631
atc gcg gcg gcc gcc aag cct ccc ctg ggc	agt ccg ccg ccg aaa acg acg gta
ile ala ala ala lys pro pro leu gly	ser pro pro pro lys thr thr val

SEQ ID N° 49A (suite 1)

FIGURE 49A (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

168/185

1/1
 cat cta cac cgc tca aca gcc ggg cca gac gct gcc ggt cgg tgc tgc cga gaa ggc ggt
 his leu his arg ser thr ala gly pro asp ala ala gly arg cys cys arg glu gly gly
 61/21
 gat ccg tgg cga gtt gtt cat gtc gcg gcg cac cac cgc cga cca acg ggt gct tgc cat
 asp pro trp arg val val his val ala ala his his arg arg pro thr gly ala cys his
 121/41
 ccg tct gac caa cgg tag ttc gct gct gat ctc caa aag tct caa gcc cac cga agc agt
 pro ser asp gln arg AMB phe ala ala asp leu gln lys ser gln ala his arg ser ser
 181/61
 cat gaa caa gct gcg ttg ggt gct att gat cgt ggg tgg gat cgg ggt ggc ggt cgc cgc
 his glu gln ala ala leu gly ala ile asp arg gly trp asp arg gly gly gly arg arg
 241/81
 ggt ggc cgg ggg gat ggt cac ccg ggc cgg gct gag gcc ggt ggg ccg cct cac cga agc
 gly gly arg gly asp gly his pro gly arg ala glu ala gly gly pro pro his arg ser
 301/101
 ggc cga gcg ggt ggc gcg aac cga cga cct gcg gcc cat ccc cgt ctt cgg cag cga cga
 gly arg ala gly gly ala asn arg arg pro ala ala his pro arg leu arg gln arg arg
 361/121
 att ggc cag gct gac aga ggc att caa ttt aat gct gcg ggc gct ggc cga gtc acg gga
 ile gly gln ala asp arg gly ile gln phe asn ala ala gly ala gly arg val thr gly
 421/141
 acg gca ggc aag gct ggt tac cga cgc cgg aca tga att gcg tac ccc gct aac gtc gct
 thr ala gly lys ala gly tyr arg arg arg thr OPA ile ala tyr pro ala asn val ala
 481/161
 gcg cac caa tgt cga act ctt gat ggc ctc gat gcc ccc ggg ggc tcc gcg gct acc caa
 ala his gln cys arg thr leu asp gly leu asp gly pro gly gly ser ala ala thr gln
 541/181
 gca gga gat ggt cga cct gcg tgc cga tgt gct ggc tca aat cga gga att gtc cac act
 ala gly asp gly arg pro ala cys arg cys ala gly ser asn arg gly ile val his thr
 601/201
 ggt agg cga ttt ggt gga cct gtc ccg agg cga cgc cgg aga agt ggt gca cga gcc ggt
 gly arg arg phe gly gly pro val pro arg arg arg arg arg ser gly ala arg ala gly
 661/221
 cga cat ggc tga cgt cgt cga ccg cag cct gga gcg ggt cag gcg gcg gcg caa cga tat
 arg his gly OPA arg arg arg pro gln pro gly ala gly gln ala ala ala gln arg tyr
 721/241
 cct ttt cga cgt cga ggt gat tgg gtg gca ggt tta tgg cga tac cgc tgg att gtc gcg
 pro phe arg arg arg gly asp trp val ala gly leu trp arg tyr arg trp ile val ala
 781/261
 gat ggc gct taa cct gat gga caa cgc cgc gaa gtg gag ccc gcc ggg cgg cca cgt ggg
 asp gly ala OCH pro asp gly gln arg arg glu val glu pro ala gly arg pro arg gly
 841/281
 tgt cag gct gag cca gct cga cgc gtc gca cgc tga gct ggt ggt ttc cga ccg cgg ccc
 cys gln ala glu pro ala arg arg val ala arg OPA ala gly gly phe arg pro arg pro
 901/301
 ggg cat tcc cgt gca gga gcg ccg tct ggt gtt tga acg gtt tta ccg gtc ggc atc ggc
 gly his ser arg ala gly ala pro ser gly val OPA thr val leu pro val gly ile gly
 961/321
 acg ggc gtt gcc ggg ttc ggg cct ccg gtt ggc gat cgt caa aca ggt ggt gct caa cca
 thr gly val ala gly phe gly pro arg val gly asp arg gln thr gly gly ala gln pro
 1021/341
 cgg cgg att gct gcg cat cga aga cac cga ccc agg cgg cca gcc ccc tgg aac gtc gat
 arg arg ile ala ala his arg arg his arg pro arg arg pro ala pro trp asn val asp

SEQ ID N° 49B

FEUILLE DE REMPLACEMENT (REGLE 26)

169/185

1081/361
 tta cgt gct gct ccc cgg ccg tcg gat gcc gat tcc gca gct tcc cgg tgc gac ggc tgg
 leu arg ala ala pro arg pro ser asp ala asp ser ala ala ser arg cys asp gly trp
 1141/381
 cgc tcg gag cac gga cat cga gaa ctc tcg ggg ttc ggc gaa cgt tat ctc agt gga atc
 arg ser glu his gly his arg glu leu ser gly phe gly glu arg tyr leu ser gly ile
 1201/401
 tca gtc cac gcg cgc aac cta gtt gtg cag tta ctg ttg aaa gcc aca ccc atg cca gtc
 ser val his ala arg asn leu val val gln leu leu leu lys ala thr pro met pro val
 1261/421
 cac gca tgg cca agt tgg ccc gag tag tgg gcc tag tac agg aag agc aac cta gcg aca
 his ala trp pro ser trp pro glu AMB trp ala AMB tyr arg lys ser asn leu ala thr
 1321/441
 tga cga atc acc cac ggt att cgc cac cgc cgc agc agc cgg gaa ccc cag gtt atg ctc
 OPA arg ile thr his gly ile arg his arg arg ser ser arg glu pro gln val met leu
 1381/461
 agg ggc agc agc aaa cgt aca gcc agc agt tcg act ggc gtt acc cac cgt ccc cgc ccc
 arg gly ser ser lys arg thr ala ser ser ser thr gly val thr his arg pro arg pro
 1441/481
 cgc agc caa ccc agt acc gtc aac cct acg agg cgt tgg gtg gta ccc ggc cgg gtc tga
 arg ser gln pro ser thr val asn pro thr arg arg trp val val pro gly arg val OPA
 1501/501
 tac ctg gcg tga ttc cga cca tga cgc ccc ctc ctg gga tgg ttc gcc aac gcc ctc gtg
 tyr leu ala OPA phe arg pro OPA arg pro leu leu gly trp phe ala asn ala leu val
 1561/521
 cag gca tgt tgg cca tcg gcg cgg tga cga tag cgg tgg tgt ccg ccg gca tcg gcg gcg
 gln ala cys trp pro ser ala arg OPA arg AMB arg trp cys pro pro ala ser ala ala
 1621/541
 cgg ccg cat ccc tgg tcg ggt tca acc ggg cac ccg ccg gcc cca gcg gcg gcc cag tgg
 arg pro his pro trp ser gly ser thr gly his pro pro ala pro ala ala ala gln trp
 1681/561
 ctg cca gcg cgg cgc caa gca tcc ccg cag caa aca tgc cgc cgg ggt cgg tcg aac agg
 leu pro ala arg arg gln ala ser pro gln gln thr cys arg arg gly arg ser asn arg
 1741/581
 tgg cgg cca agg tgg tgc cca gtg tcg tca tgt tgg aaa ccg atc tgg gcc gcc agt cgg
 trp arg pro arg trp cys pro val ser ser cys trp lys pro ile trp ala ala ser arg
 1801/601
 agg agg gct ccg gca tca ttc tgt ctg ccg agg ggc tga tct tga cca aca acc acg tga
 arg arg ala pro ala ser phe cys leu pro arg gly OPA ser OPA pro thr thr thr OPA
 1861/621
 tcg cgg cgg ccg cca agc ctc ccc tgg gca gtc cgc cgc cga aaa cga cgg ta
 ser arg arg pro pro ser leu pro trp ala val arg arg arg lys arg arg

SEQ ID N° 49B (suite 1)

FIGURE 49B (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

170/185

```

1/1                                31/11
atc tac acc gct caa cag ccg ggc cag acg ctg ccg gtc ggt gct gcc gag aag gcg gtg
ile tyr thr ala gln gln pro gly gln thr leu pro val gly ala ala glu lys ala val
61/21                                91/31
atc cgt ggc gag ttg ttc atg tcg cgg cgc acc acc gcc gac caa cgg gtg ctt gcc atc
ile arg gly glu leu phe met ser arg arg thr thr ala asp gln arg val leu ala ile
121/41                                151/51
cgt ctg acc aac ggt agt tcg ctg ctg atc tcc aaa agt ctc aag ccc acc gaa gca gtc
arg leu thr asn gly ser ser leu leu ile ser lys ser leu lys pro thr glu ala val
181/61                                211/71
atg aac aag ctg cgt tgg gtg cta ttg atc gtg ggt ggg atc ggg gtg gcg gtc gcc gcg
met asn lys leu arg trp val leu leu ile val gly gly ile gly val ala val ala ala
241/81                                271/91
gtg gcc ggg ggg atg gtc acc cgg gcc ggg ctg agg ccg gtg ggc cgc ctc acc gaa gcg
val ala gly gly met val thr arg ala gly leu arg pro val gly arg leu thr glu ala
301/101                               331/111
gcc gag cgg gtg gcg cga acc gac gac ctg cgg ccc atc ccc gtc ttc ggc agc gac gaa
ala glu arg val ala arg thr asp asp leu arg pro ile pro val phe gly ser asp glu
361/121                               391/131
ttg gcc agg ctg aca gag gca ttc aat tta atg ctg ccg gcg ctg gcc gag tca cgg gaa
leu ala arg leu thr glu ala phe asn leu met leu arg ala leu ala glu ser arg glu
421/141                               451/151
cgg cag gca agg ctg gtt acc gac gcc gga cat gaa ttg cgt acc ccg cta acg tcg ctg
arg gln ala arg leu val thr asp ala gly his glu leu arg thr pro leu thr ser leu
481/161                               511/171
cgc acc aat gtc gaa ctc ttg atg gcc tcg atg gcc ccg ggg gct ccg cgg cta ccc aag
arg thr asn val glu leu leu met ala ser met ala pro gly ala pro arg leu pro lys
541/181                               571/191
cag gag atg gtc gac ctg cgt gcc gat gtg ctg gct caa atc gag gaa ttg tcc aca ctg
gln glu met val asp leu arg ala asp val leu ala gln ile glu glu leu ser thr leu
601/201                               631/211
gta ggc gat ttg gtg gac ctg tcc cga ggc gac gcc gga gaa gtg gtg cac gag ccg gtc
val gly asp leu val asp leu ser arg gly asp ala gly glu val val his glu pro val
661/221                               691/231
gac atg gct gac gtc gtc gac cgc agc ctg gag cgg gtc agg ccg cgg cgc aac gat atc
asp met ala asp val val asp arg ser leu glu arg val arg arg arg arg asn asp ile
721/241                               751/251
ctt ttc gac gtc gag gtg att ggg tgg cag gtt tat ggc gat acc gct gga ttg tcg ccg
leu phe asp val glu val ile gly trp gln val tyr gly asp thr ala gly leu ser arg
781/261                               811/271
atg gcg ctt aac ctg atg gac aac gcc gcg aag tgg agc ccg ccg ggc ggc cac gtg ggt
met ala leu asn leu met asp asn ala ala lys trp ser pro pro gly gly his val gly
841/281                               871/291
gtc agg ctg agc cag ctc gac gcg tcg cac gct gag ctg gtg gtt tcc gac cgc ggc ccg
val arg leu ser gln leu asp ala ser his ala glu leu val val ser asp arg gly pro
901/301                               931/311
ggc att ccc gtg cag gag cgc cgt ctg gtg ttt gaa ccg ttt tac ccg tcg gca tcg gca
gly ile pro val gln glu arg arg leu val phe glu arg phe tyr arg ser ala ser ala
961/321                               991/331
cgg gcg ttg ccg ggt tcg ggc ctc ggg ttg gcg atc gtc aaa cag gtg gtg ctc aac cac
arg ala leu pro gly ser gly leu gly leu ala ile val lys gln val val leu asn his

```

SEQ ID N° 49C

FIGURE 49C
FEUILLE DE REMPLACEMENT (REGLE 26)

171/185

1021/341
 ggc gga ttg ctg cgc atc gaa gac acc gac cca ggc ggc cag ccc cct gga acg tcg att
 gly gly leu leu arg ile glu asp thr asp pro gly gly gln pro pro gly thr ser ile
 1081/361
 tac gtg ctg ctc ccc ggc cgt cgg atg ccg att ccg cag ctt ccc ggt gcg acg gct ggc
 tyr val leu leu pro gly arg arg met pro ile pro gln leu pro gly ala thr ala gly
 1141/381
 gct cgg agc acg gac atc gag aac tct ccg ggt tcg gcg aac gtt atc tca gtg gaa tct
 ala arg ser thr asp ile glu asn ser arg gly ser ala asn val ile ser val glu ser
 1201/401
 cag tcc acg cgc gca acc tag ttg tgc agt tac tgt tga aag cca cac cca tgc cag tcc
 gln ser thr arg ala thr AMB leu cys ser tyr cys OPA lys pro his pro cys gln ser
 1261/421
 acg cat ggc caa gtt ggc ccg agt agt ggg cct agt aca gga aga gca acc tag cga cat
 thr his gly gln val gly pro ser ser gly pro ser thr gly arg ala thr AMB arg his
 1321/441
 gac gaa tca ccc acg gta ttc gcc acc gcc gca gca gcc ggg aac ccc agg tta tgc tca
 asp glu ser pro thr val phe ala thr ala ala ala ala gly asn pro arg leu cys ser
 1381/461
 ggg gca gca gca aac gta cag cca gca gtt cga ctg gcg tta ccc acc gtc ccc gcc ccc
 gly ala ala ala asn val gln pro ala val arg leu ala leu pro thr val pro ala pro
 1441/481
 gca gcc aac cca gta ccg tca acc cta cga ggc gtt ggg tgg tac ccg gcc ggg tct gat
 ala ala asn pro val pro ser thr leu arg gly val gly trp tyr pro ala gly ser asp
 1501/501
 acc tgg cgt gat tcc gac cat gac gcc ccc tcc tgg gat ggt tcg cca acg ccc tcg tgc
 thr trp arg asp ser asp his asp ala pro ser trp asp gly ser pro thr pro ser cys
 1561/521
 agg cat gtt ggc cat ccg cgc ggt gac gat agc ggt ggt gtc cgc ccg cat ccg ccg cgc
 arg his val gly his arg arg gly asp asp ser gly gly val arg arg his arg arg arg
 1621/541
 ggc cgc atc cct ggt ccg gtt caa ccg ggc acc cgc ccg ccc cag ccg ccg ccc agt ggc
 gly arg ile pro gly arg val gln pro gly thr arg arg pro gln arg arg pro ser gly
 1681/561
 tgc cag cgc ggc gcc aag cat ccc cgc agc aaa cat gcc gcc ggg gtc ggt cga aca ggt
 cys gln arg gly ala lys his pro arg ser lys his ala ala gly val gly arg thr gly
 1741/581
 ggc ggc caa ggt ggt gcc cag tgt cgt cat gtt gga aac cga tct ggg ccg cca gtc gga
 gly gly gln gly gly ala gln cys arg his val gly asn arg ser gly pro pro val gly
 1801/601
 gga ggg ctc ccg cat cat tct gtc tgc cga ggg gct gat ctt gac caa caa cca cgt gat
 gly gly leu arg his his ser val cys arg gly ala asp leu asp gln gln pro arg asp
 1861/621
 cgc ggc ggc cgc caa gcc tcc cct ggg cag tcc gcc gcc gaa aac gac ggt a
 arg gly gly arg gln ala ser pro gly gln ser ala ala glu asn asp gly

SEQ ID N° 49C (suite 1)

FIGURE 49C (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

172/185

Séquence codante Rv0983 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq60A:

```

1/1                               31/11
atg gcc aag ttg gcc cga gta gtg ggc cta gta cag gaa gag caa cct agc gac atg acg
Met ala lys leu ala arg val val gly leu val gln glu glu gln pro ser asp met thr
61/21                               91/31
aat cac cca cgg tat tcg cca ccg ccg cag cag ccg gga acc cca ggt tat gct cag ggg
asn his pro arg tyr ser pro pro pro gln gln pro gly thr pro gly tyr ala gln gly
121/41                               151/51
cag cag caa acg tac agc cag cag ttc gac tgg cgt tac cca ccg tcc ccg ccc ccg cag
gln gln gln thr tyr ser gln gln phe asp trp arg tyr pro pro ser pro pro pro gln
181/61                               211/71
cca acc cag tac cgt caa ccc tac gag gcg ttg ggt ggt acc ccg ccg ggt ctg ata cct
pro thr gln tyr arg gln pro tyr glu ala leu gly gly thr arg pro gly leu ile pro
241/81                               271/91
ggc gtg att ccg acc atg acg ccc cct cct ggg atg gtt cgc caa cgc cct cgt gca ggc
gly val ile pro thr met thr pro pro pro gly met val arg gln arg pro arg ala gly
301/101                               331/111
atg ttg gcc atc ggc gcg gtg acg ata gcg gtg gtg tcc gcc ggc atc ggc ggc gcg gcc
met leu ala ile gly ala val thr ile ala val val ser ala gly ile gly gly ala ala
361/121                               391/131
gca tcc ctg gtc ggg ttc aac ccg gca ccc gcc gcc ccc agc ggc ggc cca gtg gct gcc
ala ser leu val gly phe asn arg ala pro ala gly pro ser gly gly pro val ala ala
421/141                               451/151
agc gcg gcg cca agc atc ccc gca gca aac atg ccg ccg ggg tcg gtc gaa cag gtg gcg
ser ala ala pro ser ile pro ala ala asn met pro pro gly ser val glu gln val ala
481/161                               511/171
gcc aag gtg gtg ccc agt gtc gtc atg ttg gaa acc gat ctg ggc cgc cag tcg gag gag
ala lys val val pro ser val val met leu glu thr asp leu gly arg gln ser glu glu
541/181                               571/191
ggc tcc ggc atc att ctg tct gcc gag ggg ctg atc ttg acc aac aac cac gtg atc gcg
gly ser gly ile ile leu ser ala glu gly leu ile leu thr asn asn his val ile ala
601/201                               631/211
gcg gcc gcc aag cct ccc ctg ggc agt ccg ccg ccg aaa acg acg gta acc ttc tct gac
ala ala ala lys pro pro leu gly ser pro pro pro lys thr thr val thr phe ser asp
661/221                               691/231
ggg ccg acc gca ccc ttc acg gtg gtg ggg gct gac ccc acc agt gat atc gcc gtc gtc
gly arg thr ala pro phe thr val val gly ala asp pro thr ser asp ile ala val val
721/241                               751/251
cgt gtt cag ggc gtc tcc ggg ctc acc ccg atc tcc ctg ggt tcc tcc tcg gac ctg agg
arg val gln gly val ser gly leu thr pro ile ser leu gly ser ser ser asp leu arg
781/261                               811/271
gtc ggt cag ccg gtg ctg gcg atc ggg tcg ccg ctc ggt ttg gag ggc acc gtg acc acg
val gly gln pro val leu ala ile gly ser pro leu gly leu glu gly thr val thr thr

```

SEQ ID N° 49D

FIGURE 49D

FEUILLE DE REMPLACEMENT (REGLE 26)

173/185

841/281

ggg atc gtc agc gct ctc aac cgt cca gtg tgc acg acc ggc gag gcc ggc aac cag aac
gly ile val ser ala leu asn arg pro val ser thr thr gly glu ala gly asn gln asn
901/301

acc gtg ctg gac gcc att cag acc gac gcc gcg atc aac ccc ggt aac tcc ggg ggc gcg
thr val leu asp ala ile gln thr asp ala ala ile asn pro gly asn ser gly gly ala
961/321

ctg gtg aac atg aac gct caa ctc gtc gga gtc aac tgc gcc att gcc acg ctg ggc gcg
leu val asn met asn ala gln leu val gly val asn ser ala ile ala thr leu gly ala
1021/341

gac tca gcc gat gcg cag agc ggc tgc atc ggt ctc ggt ttt gcg att cca gtc gac cag
asp ser ala asp ala gln ser gly ser ile gly leu gly phe ala ile pro val asp gln
1081/361

gcc aag cgc atc gcc gac gag ttg atc agc acc ggc aag gcg tca cat gcc tcc ctg ggt
ala lys arg ile ala asp glu leu ile ser thr gly lys ala ser his ala ser leu gly
1141/381

gtg cag gtg acc aat gac aaa gac acc ctg ggc gcc aag atc gtc gaa gta gtg gcc ggt
val gln val thr asn asp lys asp thr leu gly ala lys ile val glu val val ala gly
1201/401

ggt gct gcc gcg aac gct gga gtg ccg aag ggc gtc gtt gtc acc aag gtc gac gac cgc
gly ala ala ala asn ala gly val pro lys gly val val val thr lys val asp asp arg
1261/421

ccg atc aac agc gcg gac gcg ttg gtt gcc gcc gtg ccg tcc aaa gcg ccg ggc gcc acg
pro ile asn ser ala asp ala leu val ala ala val arg ser lys ala pro gly ala thr
1321/441

gtg gcg cta acc ttt cag gat ccc tgc ggc ggt agc cgc aca gtg caa gtc acc ctc ggc
val ala leu thr phe gln asp pro ser gly gly ser arg thr val gln val thr leu gly
1381/461

aag gcg gag cag tga
lys ala glu gln OPA

871/291

931/311

991/331

1051/351

1111/371

1171/391

1231/411

1291/431

1351/451

SEQ ID N° 49D (suite 1)

FIGURE 49D (suite 1)

174/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0983

```

1/1                               31/11
tga gcc agc tcg acg cgt cgc acg ctg agc tgg tgg ttt ccg acc gcg gcc cgg gca ttc
OPA ala ser ser thr arg arg thr leu ser trp trp phe pro thr ala ala arg ala phe
61/21                               91/31
ccg tgc agg agc gcc gtc tgg tgt ttg aac ggt ttt acc ggt cgg cat cgg cac ggg cgt
pro cys arg ser ala val trp cys leu asn gly phe thr gly arg his arg his gly arg
121/41                               151/51
tgc cgg gtt cgg gcc tcg ggt tgg cga tcg tca aac agg tgg tgc tca acc acg gcg gat
cys arg val arg ala ser gly trp arg ser ser asn arg trp cys ser thr thr ala asp
181/61                               211/71
tgc tgc gca tcg aag aca ccg acc cag gcg gcc agc ccc ctg gaa cgt cga ttt acg tgc
cys cys ala ser lys thr pro thr gln ala ala ser pro leu glu arg arg phe thr cys
241/81                               271/91
tgc tcc ccg gcc gtc gga tgc cga ttc cgc agc ttc ccg gtg cga cgg ctg gcg ctc gga
cys ser pro ala val gly cys arg phe arg ser phe pro val arg arg leu ala leu gly
301/101                               331/111
gca cgg aca tcg aga act ctc ggg gtt cgg cga acg tta tct cag tgg aat ctc agt cca
ala arg thr ser arg thr leu gly val arg arg thr leu ser gln trp asn leu ser pro
361/121                               391/131
cgc gcg caa cct agt tgt gca gtt act gtt gaa agc cac acc cat gcc agt cca cgc atg
arg ala gln pro ser cys ala val thr val glu ser his thr his ala ser pro arg met
421/141                               451/151
gcc aag ttg gcc cga gta gtg ggc cta gta cag gaa gag caa cct agc gac atg acg aat
ala lys leu ala arg val val gly leu val gln glu glu gln pro ser asp met thr asn
481/161                               511/171
cac cca cgg tat tcg cca ccg ccg cag cag ccg gga acc cca ggt tat gct cag ggg cag
his pro arg tyr ser pro pro pro gln gln pro gly thr pro gly tyr ala gln gly gln
541/181                               571/191
cag caa acg tac agc cag cag ttc gac tgg cgt tac cca ccg tcc ccg ccc ccg cag cca
gln gln thr tyr ser gln gln phe asp trp arg tyr pro pro ser pro pro pro gln pro
601/201                               631/211
acc cag tac cgt caa ccc tac gag gcg ttg ggt ggt acc ccg ccg ggt ctg ata cct ggc
thr gln tyr arg gln pro tyr glu ala leu gly gly thr arg pro gly leu ile pro gly
661/221                               691/231
gtg att ccg acc atg acg ccc cct cct ggg atg gtt cgc caa cgc cct cgt gca ggc atg
val ile pro thr met thr pro pro pro gly met val arg gln arg pro arg ala gly met
721/241                               751/251
ttg gcc atc ggc gcg gtg acg ata gcg gtg gtg tcc gcc ggc atc ggc ggc gcg gcc gca
leu ala ile gly ala val thr ile ala val val ser ala gly ile gly gly ala ala ala
781/261                               811/271
tcc ctg gtc ggg ttc aac ccg gca ccc gcc ggc ccc agc ggc ggc cca gtg gct gcc agc
ser leu val gly phe asn arg ala pro ala gly pro ser gly gly pro val ala ala ser
841/281                               871/291
gcg gcg cca agc atc ccc gca gca aac atg ccg ccg ggg tcg gtc gaa cag gtg gcg gcc
ala ala pro ser ile pro ala ala asn met pro pro gly ser val glu gln val ala ala

```

SEQ ID N° 49F

FIGURE 49F

FEUILLE DE REMPLACEMENT (REGLE 26)

175/185

901/301	931/311
aag gtg gtg ccc agt gtc gtc atg ttg gaa	acc gat ctg ggc cgc cag tcg gag gag ggc
lys val val pro ser val val met leu glu	thr asp leu gly arg gln ser glu glu gly
961/321	991/331
tcc ggc atc att ctg tct gcc gag ggg ctg	atc ttg acc aac aac cac gtg atc gcg gcg
ser gly ile ile leu ser ala glu gly leu	ile leu thr asn asn his val ile ala ala
1021/341	1051/351
gcc gcc aag cct ccc ctg ggc agt ccg ccg	ccg aaa acg acg gta acc ttc tct gac ggg
ala ala lys pro pro leu gly ser pro pro	pro lys thr thr val thr phe ser asp gly
1081/361	1111/371
cgg acc gca ccc ttc acg gtg gtg ggg gct	gac ccc acc agt gat atc gcc gtc gtc cgt
arg thr ala pro phe thr val val gly ala	asp pro thr ser asp ile ala val val arg
1141/381	1171/391
gtt cag ggc gtc tcc ggg ctc acc ccg atc	tcc ctg ggt tcc tcc tcg gac ctg agg gtc
val gln gly val ser gly leu thr pro ile	ser leu gly ser ser ser asp leu arg val
1201/401	1231/411
ggt cag ccg gtg ctg gcg atc ggg tcg ccg	ctc ggt ttg gag ggc acc gtg acc acg ggg
gly gln pro val leu ala ile gly ser pro	leu gly leu glu gly thr val thr thr gly
1261/421	1291/431
atc gtc agc gct ctc aac cgt cca gtg tcg	acg acc ggc gag gcc ggc aac cag aac acc
ile val ser ala leu asn arg pro val ser	thr thr gly glu ala gly asn gln asn thr
1321/441	1351/451
gtg ctg gac gcc att cag acc gac gcc gcg	atc aac ccc ggt aac tcc ggg ggc gcg ctg
val leu asp ala ile gln thr asp ala ala	ile asn pro gly asn ser gly gly ala leu
1381/461	1411/471
gtg aac atg aac gct caa ctc gtc gga gtc	aac tcg gcc att gcc acg ctg ggc gcg gac
val asn met asn ala gln leu val gly val	asn ser ala ile ala thr leu gly ala asp
1441/481	1471/491
tca gcc gat gcg cag agc ggc tcg atc ggt	ctc ggt ttt gcg att cca gtc gac cag gcc
ser ala asp ala gln ser gly ser ile gly	leu gly phe ala ile pro val asp gln ala
1501/501	1531/511
aag cgc atc gcc gac gag ttg atc agc acc	ggc aag gcg tca cat gcc tcc ctg ggt gtg
lys arg ile ala asp glu leu ile ser thr	gly lys ala ser his ala ser leu gly val
1561/521	1591/531
cag gtg acc aat gac aaa gac acc ctg ggc	gcc aag atc gtc gaa gta gtg gcc ggt ggt
gln val thr asn asp lys asp thr leu gly	ala lys ile val glu val val ala gly gly
1621/541	1651/551
gct gcc gcg aac gct gga gtg ccg aag ggc	gtc gtt gtc acc aag gtc gac gac cgc ccg
ala ala ala asn ala gly val pro lys gly	val val val thr lys val asp asp arg pro
1681/561	1711/571
atc aac agc gcg gac gcg ttg gtt gcc gcc	gtg cgg tcc aaa gcg ccg ggc gcc acg gtg
ile asn ser ala asp ala leu val ala ala	val arg ser lys ala pro gly ala thr val
1741/581	1771/591
gcg cta acc ttt cag gat ccc tcg ggc ggt	agc cgc aca gtg caa gtc acc ctc ggc aag
ala leu thr phe gln asp pro ser gly gly	ser arg thr val gln val thr leu gly lys
1801/601	
gcg gag cag tga	
ala glu gln OPA	

SEQ ID N° 49F (suite 1)

FIGURE 49F (suite 1)

FEUILLE DE REMPLACEMENT (REGLE 26)

176/185

Fragment amplifié par PCR d'après les similarités de séquence avec une sérine protéase de la famille HtrA de E. coli (création du site SnaBI à l'extrémité 3') et sous cloné dans le vecteur pJVEDa:

```

1/1                               31/11
gat ccg gcg ggg cgg gtg tcg gcg cag gcg tgg ctg gcg gtc acg gcg gtg cgg gcg gtg
asp pro ala gly arg val ser ala gln ala trp leu ala val thr ala val arg ala val
61/21                               91/31
ccg ccg ggc tgt ggg gcg ccg gcg gcg gcg gtg gca atg gcg gga acg gcg ccg atg cca
pro pro gly cys gly ala pro ala ala ala val ala met ala gly thr ala pro met pro
121/41                               151/51
aca tcg tca gcg gtg gag acg gtg gcc tcg gcg gtg ccg gtg gcg gtg gcg gat ggc tct
thr ser ser ala val glu thr val ala ser ala val pro val ala val ala asp gly ser
181/61                               211/71
acg gcg acg gcg ggg ccg gcg gac acg gcg gac aag gcg caa tcg gcc tcg gcg gcg gcg
thr ala thr ala gly pro ala asp thr ala asp lys ala gln ser ala ser ala ala ala
241/81                               271/91
ccg gcg gcg acg ggg gcc agg gcg gcg ccg gcc gcg gac tgt ggg gta ctg gcg gcg ccg
pro ala ala thr gly ala arg ala ala pro ala ala asp cys gly val leu ala ala pro
301/101                               331/111
gcg gac acg gcg ggc aag gcg gtg gta ccg ggg gcc cac cgc tgc ccg gtc agg cag gca
ala asp thr ala gly lys ala val val pro gly ala his arg cys pro val arg gln ala
361/121                               391/131
tgg gcg ccg cgg gtg gcg ccg gtg ggc tga tcg gca acg gcg ggg ccg gcg gcg acg gcg
trp ala pro arg val ala pro val gly OPA ser ala thr ala gly pro ala ala thr ala
421/141                               451/151
gtg tcg gcg cgt ccg gcg ggg tcg ccg gag tag gcg gtg ccg gcg gga acg cca tgc tga
val ser ala arg pro ala gly ser pro glu AMB ala val pro ala gly thr pro cys OPA
481/161                               511/171
tcg ggc acg gcg gcg ccg gcg gcg ccg gcg gag aca gca gtt tcg cta atg gcg cgg ccg
ser gly thr ala ala pro ala ala pro ala glu thr ala val ser leu met ala arg pro
541/181                               571/191
gcg gcg ccg gcg gtg ccg gag ggc acc tct tcg gca atg gcg ggt ccg gcg gcc acg gcg
ala ala arg ala val pro glu gly thr ser ser ala met ala gly pro ala ala thr ala
601/201                               631/211
gag ccg tca ccg ccg gca aca ccg gta tcg gtg gcg ccg gcg gcg tcg gtg ggg acg cca
glu pro ser arg pro ala thr pro val ser val ala pro ala ala ser val gly thr pro
661/221                               691/231
ggc tga tcg gcc acg gtg gcg ccg gcg gtg ccg gcg ggg acc gcg ccg gag cct tgg ttg
gly OPA ser ala thr val ala pro ala val pro ala gly thr ala pro glu pro trp leu
721/241                               751/251
gcc gtg acg gcg ggc ccg gtg gga acg ggg gcg ctg gcg gcc agc tat acg gca acg gcg
ala val thr ala gly pro val gly thr gly ala leu ala ala ser tyr thr ala thr ala
781/261                               811/271
gcg acg gcg ccc ccg gca ccg gcg gaa cac tgc agg ccg ccg tga gcg gat tgg tga ccg
ala thr ala pro pro ala pro ala glu his cys arg arg arg OPA ala asp trp OPA arg
841/281                               871/291
ctt tgt tcg gtg cac ccg gcc aac ccg gcg aca ccg gcc aac ccg gct agc ccc gat caa
leu cys ser val his pro ala asn pro ala thr pro ala asn pro ala ser pro asp gln
901/301                               931/311
cga ggg ttt ccg tgc ccg tcc ggg gca tgg cca tcc gct gag ctg gcg atc tgg act acg
arg gly phe arg cys arg ser gly ala trp pro ser ala glu leu ala ile trp thr thr
961/321                               991/331
ttg gtg tag aaa aat cct gcc gcc ccg acc ctt aag gct ggg aca att tct gat agc tac
leu val AMB lys asn pro ala ala arg thr leu lys ala gly thr ile ser asp ser tyr
1021/341                               1051/351
ccc gac aca gga ggt tac ggg atg agc aat tcg cgc cgc cgc tca ctc agg tgg tca tgg
pro asp thr gly gly tyr gly met ser asn ser arg arg arg ser leu arg trp ser trp
1081/361                               1111/371
ttg ctg agc gtg ctg gct gcc gtc ggg ctg ggc ctg gcc acg gcg ccg gcc cag gcg gcc
leu leu ser val leu ala ala val gly leu gly leu ala thr ala pro ala gln ala ala
1141/381
ccg ccg gcc ttg tcg cag gac ccg tt
pro pro ala leu ser gln asp arg

```

SEQ ID N° 50A

FEUILLE DE REMPLACEMENT (REGLE 26)

FIGURE 50A

177/185

1/1
 atc cgg cgg ggc ggg tgt cgg cgc agg cgt ggc tgg cgg tca cgg cgg tgc ggg cgg tgc
 ile arg arg gly gly cys arg arg arg arg gly trp arg ser arg arg cys gly arg cys
 61/21
 cgc cgg gct gtg ggg cgc cgg cgg cgg cgg tgg caa tgg cgg gaa cgg cgc cga tgc caa
 arg arg ala val gly arg arg arg arg trp gln trp arg glu arg arg arg cys gln
 121/41
 cat cgt cag cgg tgg aga cgg tgg cct cgg cgg tgc cgg tgg cgg tgg cgg atg gct cta
 his arg gln arg trp arg arg trp pro arg arg cys arg trp arg trp arg met ala leu
 181/61
 cgg cga cgg cgg ggc cgg cgg aca cgg cgg aca agg cgc aat cgg cct cgg cgg cgg cgc
 arg arg arg arg gly arg arg thr arg arg thr arg arg asn arg pro arg arg arg arg
 241/81
 cgg cgg cga cgg ggg cca ggg cgg cgc cgg ccg cgg act gtg ggg tac tgg cgg cgc cgg
 arg arg arg arg gly pro gly arg arg arg pro arg thr val gly tyr trp arg arg arg
 301/101
 cgg aca cgg cgg gca agg cgg tgg tac cgg ggg ccc acc gct gcc cgg tca ggc agg cat
 arg thr arg arg ala arg arg trp tyr arg gly pro thr ala ala arg ser gly arg his
 361/121
 ggg cgc cgc ggg tgg cgc cgg tgg gct gat cgg caa cgg cgg ggc cgg cgg cga cgg cgg
 gly arg arg gly trp arg arg trp ala asp arg gln arg arg gly arg arg arg arg arg
 421/141
 tgt cgg cgc gtc cgg cgg ggt cgc cgg agt agg cgg tgc cgg cgg gaa cgc cat gct gat
 cys arg arg val arg arg gly arg arg ser arg arg cys arg arg glu arg his ala asp
 481/161
 cgg gca cgg cgg cgc cgg cgg cgc cgg cgg aga cag cag ttt cgc taa tgg cgc ggc cgg
 arg ala arg arg arg arg arg arg arg arg arg gln gln phe arg OCH trp arg gly arg
 541/181
 cgg cgc ggg cgg tgc cgg agg gca cct ctt cgg caa tgg cgg gtc cgg cgg cca cgg cgg
 arg arg gly arg cys arg arg ala pro leu arg gln trp arg val arg arg pro arg arg
 601/201
 agc cgt cac ggc cgg caa cac cgg tat cgg tgg cgc cgg cgg cgt cgg tgg gga cgc cag
 ser arg his gly arg gln his arg tyr arg trp arg arg arg arg trp gly arg gln
 661/221
 gct gat cgg cca cgg tgg cgc cgg cgg tgc cgg cgg gga ccg cgc cgg agc ctt ggt tgg
 ala asp arg pro arg trp arg arg arg cys arg arg gly pro arg arg ser leu gly trp
 721/241
 ccg tga cgg cgg gcc cgg tgg gaa cgg ggg cgc tgg cgg cca gct ata cgg caa cgg cgg
 pro OPA arg arg ala arg trp glu arg gly arg trp arg pro ala ile arg gln arg arg
 781/261
 cga cgg cgc ccc cgg cac cgg cgg aac act gca ggc ggc ggt gag cgg att ggt gac ggc
 arg arg arg pro arg his arg arg asn thr ala gly gly gly glu arg ile gly asp gly
 841/281
 ttt gtt cgg tgc acc cgg cca acc cgg cga cac cgg cca acc cgg cta gcc ccg atc aac
 phe val arg cys thr arg pro thr arg arg his arg pro thr arg leu ala pro ile asn
 901/301
 gag ggt ttc ggt gcc ggt ccg ggg cat ggc cat ccg ctg agc tgg cga tct gga cta cgt
 glu gly phe gly ala gly pro gly his gly his pro leu ser trp arg ser gly leu arg
 961/321
 tgg tgt aga aaa atc ctg ccg ccc gga ccc tta agg ctg gga caa ttt ctg ata gct acc
 trp cys arg lys ile leu pro pro gly pro leu arg leu gly gln phe leu ile ala thr
 1021/341
 ccg aca cag gag gtt acg gga tga gca att cgc gcc gcc gct cac tca ggt ggt cat ggt
 pro thr gln glu val thr gly OPA ala ile arg ala ala ala his ser gly gly his gly
 1081/361
 tgc tga gcg tgc tgg ctg ccg tcg ggc tgg gcc tgg cca cgg cgc cgg ccc agg cgg ccc
 cys OPA ala cys trp leu pro ser gly trp ala trp pro arg arg arg pro arg arg pro
 1141/381
 cgc cgg cct tgt cgc agg acc ggt t
 arg arg pro cys arg arg thr gly

SEQ ID N° 50B

FIGURE 50B
 FEUILLE DE REMPLACEMENT (REGLE 26)

178/185

1/1
tcc ggc ggg gcg ggt gtc ggc gca ggc gtg gct ggc ggt cac ggc ggt gcg ggc ggt gcc
ser gly gly ala gly val gly ala gly val ala gly gly his gly gly ala gly gly ala
61/21
gcc ggg ctg tgg ggc gcc ggc ggc ggc ggt ggc aat ggc ggg aac ggc gcc gat gcc aac
ala gly leu trp gly ala gly gly gly gly gly asn gly gly asn gly ala asp ala asn
121/41
atc gtc agc ggt gga gac ggt ggc ctc ggc ggt gcc ggt ggc ggt ggc gga tgg ctc tac
ile val ser gly gly asp gly gly leu gly gly ala gly gly gly gly gly trp leu tyr
181/61
ggc gac ggc ggg gcc ggc gga cac ggc gga caa ggc gca atc ggc ctc ggc ggc ggc gcc
gly asp gly gly ala gly gly his gly gly gln gly ala ile gly leu gly gly gly ala
241/81
ggc ggc gac ggg ggc cag ggc ggc gcc ggc cgc gga ctg tgg ggt act ggc ggc gcc ggc
gly gly asp gly gly gln gly gly ala gly arg gly leu trp gly thr gly gly ala gly
301/101
gga cac ggc ggg caa ggc ggt ggt acc ggc ggc cca ccg ctg ccc ggt cag gca ggc atg
gly his gly gly gln gly gly gly thr gly gly pro pro leu pro gly gln ala gly met
361/121
ggc gcc gcg ggt ggc gcc ggt ggg ctg atc ggc aac ggc ggg gcc ggc ggc gac ggc ggt
gly ala ala gly gly ala gly gly leu ile gly asn gly gly ala gly gly asp gly gly
421/141
gtc ggc gcg tcc ggc ggg gtc gcc gga gta ggc ggt gcc ggc ggc ggc aac gcc atg ctg atc
val gly ala ser gly gly val ala gly val gly gly ala gly gly asn ala met leu ile
481/161
ggg cac ggc ggc gcc ggc ggc gcc ggc gga gac agc agt ttc gct aat ggc gcg gcc ggc
gly his gly gly ala gly gly ala gly gly asp ser ser phe ala asn gly ala ala gly
541/181
ggc gcg ggc ggt gcc gga ggg cac ctc ttc ggc aat ggc ggg tcc ggc ggc cac ggc gga
gly ala gly gly ala gly gly his leu phe gly asn gly gly ser gly gly his gly gly
601/201
gcc gtc acg gcc ggc aac acc ggt atc ggt ggc gcc ggc ggc ggc gtc ggt ggg gac gcc agg
ala val thr ala gly asn thr gly ile gly gly ala gly gly val gly gly asp ala arg
661/221
ctg atc ggc cac ggt ggc gcc ggc ggt gcc ggc ggg gac cgc gcc gga gcc ttg gtt ggc
leu ile gly his gly gly ala gly gly ala gly gly asp arg ala gly ala leu val gly
721/241
cgt gac ggc ggg ccc ggt ggg aac ggg ggc gct ggc ggc cag cta tac ggc aac ggc ggc
arg asp gly gly pro gly gly asn gly gly ala gly gly gln leu tyr gly asn gly gly
781/261
gac ggc gcc ccc ggc acc ggc gga aca ctg cag gcg gcg gtg agc gga ttg gtg acg gct
asp gly ala pro gly thr gly gly thr leu gln ala ala val ser gly leu val thr ala
841/281
ttg ttc ggt gca ccc ggc caa ccc ggc gac acc ggc caa ccc ggc tag ccc cga tca acg
leu phe gly ala pro gly gln pro gly asp thr gly gln pro gly AMB pro arg ser thr
901/301
agg gtt tcg gtg ccg gtc cgg ggc atg gcc atc cgc tga gct ggc gat ctg gac tac gtt
arg val ser val pro val arg gly met ala ile arg OPA ala gly asp leu asp tyr val
961/321
ggt gta gaa aaa tcc tgc cgc ccg gac cct taa ggc tgg gac aat ttc tga tag cta ccc
gly val glu lys ser cys arg pro asp pro OCH gly trp asp asn phe OPA AMB leu pro
1021/341
cga cac agg agg tta cgg gat gag caa ttc gcg ccg ccg ctc act cag gtg gtc atg gtt
arg his arg arg leu arg asp glu gln phe ala pro pro leu thr gln val val met val
1081/361
gct gag cgt gct ggc tgc cgt cgg gct ggg cct ggc cac ggc gcc ggc cca ggc ggc ccc
ala glu arg ala gly cys arg arg ala gly pro gly his gly ala gly pro gly gly pro
1141/381
gcc ggc ctt gtc gca gga ccg gtt
ala gly leu val ala gly pro val

SEQ ID N° 50C

FIGURE 50C
FEUILLE DE REMPLACEMENT (REGLE 26)

179/185

Séquence codante Rv0125 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq50A:

```

1/1                                     31/11
atg agc aat tcg cgc cgc cgc tca ctc agg tgg tca tgg ttg ctg agc gtg ctg gct gcc
Met ser asn ser arg arg arg ser leu arg trp ser trp leu leu ser val leu ala ala
61/21                                     91/31
gtc ggg ctg ggc ctg gcc acg gcg ccg gcc cag gcg gcc ccg ccg gcc ttg tcg cag gac
val gly leu gly leu ala thr ala pro ala gln ala ala pro pro ala leu ser gln asp
121/41                                     151/51
cgg ttc gcc gac ttc ccc gcg ctg ccc ctc gac ccg tcc gcg atg gtc gcc caa gtg ggg
arg phe ala asp phe pro ala leu pro leu asp pro ser ala met val ala gln val gly
181/61                                     211/71
cca cag gtg gtc aac atc aac acc aaa ctg ggc tac aac aac gcc gtg ggc gcc ggg acc
pro gln val val asn ile asn thr lys leu gly tyr asn asn ala val gly ala gly thr
241/81                                     271/91
ggc atc gtc atc gat ccc aac ggt gtc gtg ctg acc aac aac cac gtg atc gcg ggc gcc
gly ile val ile asp pro asn gly val val leu thr asn asn his val ile ala gly ala
301/101                                    331/111
acc gac atc aat gcg ttc agc gtc ggc tcc ggc caa acc tac ggc gtc gat gtg gtc ggg
thr asp ile asn ala phe ser val gly ser gly gln thr tyr gly val asp val val gly
361/121                                    391/131
tat gac cgc acc cag gat gtc gcg gtg ctg cag ctg cgc ggt gcc ggt ggc ctg ccg tcg
tyr asp arg thr gln asp val ala val leu gln leu arg gly ala gly gly leu pro ser
421/141                                    451/151
gcg gcg atc ggt ggc ggc gtc gcg gtt ggt gag ccc gtc gtc gcg atg ggc aac agc ggt
ala ala ile gly gly gly val ala val gly glu pro val val ala met gly asn ser gly
481/161                                    511/171
ggg cag ggc gga acg ccc cgt gcg gtg cct ggc agg gtg gtc gcg ctc ggc caa acc gtg
gly gln gly gly thr pro arg ala val pro gly arg val val ala leu gly gln thr val
541/181                                    571/191
cag gcg tcg gat tcg ctg acc ggt gcc gaa gag aca ttg aac ggg ttg atc cag ttc gat
gln ala ser asp ser leu thr gly ala glu glu thr leu asn gly leu ile gln phe asp
601/201                                    631/211
gcc gcg atc cag ccc ggt gat tcg ggc ggg ccc gtc gtc aac ggc cta gga cag gtg gtc
ala ala ile gln pro gly asp ser gly gly pro val val asn gly leu gly gln val val
661/221                                    691/231
ggt atg aac acg gcc gcg tcc gat aac ttc cag ctg tcc cag ggt ggg cag gga ttc gcc
gly met asn thr ala ala ser asp asn phe gln leu ser gln gly gly gln gly phe ala
721/241                                    751/251
att ccg atc ggg cag gcg atg gcg atc gcg ggc cag atc cga tcg ggt ggg ggg tca ccc
ile pro ile gly gln ala met ala ile ala gly gln ile arg ser gly gly gly ser pro
781/261                                    811/271
acc gtt cat atc ggg cct acc gcc ttc ctc ggc ttg ggt gtt gtc gac aac aac ggc aac
thr val his ile gly pro thr ala phe leu gly leu gly val val asp asn asn gly asn
841/281                                    871/291
ggc gca cga gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt ctc ggc atc tcc acc
gly ala arg val gln arg val val gly ser ala pro ala ala ser leu gly ile ser thr
901/301                                    931/311
ggc gac gtg atc acc gcg gtc gac ggc gct ccg atc aac tcg gcc acc gcg atg gcg gac
gly asp val ile thr ala val asp gly ala pro ile asn ser ala thr ala met ala asp
961/321                                    991/331
gcg ctt aac ggg cat cat ccc ggt gac gtc atc tcg gtg acc tgg caa acc aag tcg ggc
ala leu asn gly his his pro gly asp val ile ser val thr trp gln thr lys ser gly
1021/341                                    1051/351
ggc acg cgt aca ggg aac gtg aca ttg gcc gag gga ccc ccg gcc tga
gly thr arg thr gly asn val thr leu ala glu gly pro pro ala OPA

```

SEQ ID N° 50D

FEUILLE DE REMPLACEMENT (REGLE 26)

180/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0125:

```

1/1                                31/11
tag aaa aat cct gcc gcc cgg acc ctt aag gct ggg aca att tct gat agc tac ccc gac
AMB lys asn pro ala ala arg thr leu lys ala gly thr ile ser asp ser tyr pro asp
61/21                                91/31
aca gga ggt tac ggg atg agc aat tcg cgc cgc cgc tca ctc agg tgg tca tgg ttg ctg
thr gly gly tyr gly met ser asn ser arg arg arg ser leu arg trp ser trp leu leu
121/41                                151/51
agc gtg ctg gct gcc gtc ggg ctg gcc ctg gcc acg gcg ccg gcc cag gcg gcc ccg ccg
ser val leu ala ala val gly leu gly leu ala thr ala pro ala gln ala ala pro pro
181/61                                211/71
gcc ttg tcg cag gac cgg ttc gcc gac ttc ccc gcg ctg ccc ctc gac ccg tcc gcg atg
ala leu ser gln asp arg phe ala asp phe pro ala leu pro leu asp pro ser ala met
241/81                                271/91
gtc gcc caa gtg ggg cca cag gtg gtc aac atc aac acc aaa ctg ggc tac aac aac gcc
val ala gln val gly pro gln val val asn ile asn thr lys leu gly tyr asn asn ala
301/101                               331/111
gtg ggc gcc ggg acc ggc atc gtc atc gat ccc aac ggt gtc gtg ctg acc aac aac cac
val gly ala gly thr gly ile val ile asp pro asn gly val val leu thr asn asn his
361/121                               391/131
gtg atc gcg ggc gcc acc gac atc aat gcg ttc agc gtc ggc tcc ggc caa acc tac ggc
val ile ala gly ala thr asp ile asn ala phe ser val gly ser gly gln thr tyr gly
421/141                               451/151
gtc gat gtg gtc ggg tat gac cgc acc cag gat gtc gcg gtg ctg cag ctg cgc ggt gcc
val asp val val gly tyr asp arg thr gln asp val ala val leu gln leu arg gly ala
481/161                               511/171
ggt ggc ctg ccg tcg gcg gcg atc ggt ggc ggc gtc gcg gtt ggt gag ccc gtc gtc gcg
gly gly leu pro ser ala ala ile gly gly gly val ala val gly glu pro val val ala
541/181                               571/191
atg ggc aac agc ggt ggg cag ggc gga acg ccc cgt gcg gtg cct ggc agg gtg gtc gcg
met gly asn ser gly gly gln gly gly thr pro arg ala val pro gly arg val val ala
601/201                               631/211
ctc ggc caa acc gtg cag gcg tcg gat tcg ctg acc ggt gcc gaa gag aca ttg aac ggg
leu gly gln thr val gln ala ser asp ser leu thr gly ala glu glu thr leu asn gly
661/221                               691/231
ttg atc cag ttc gat gcc gcg atc cag ccc ggt gat tcg ggc ggg ccg gtc gtc aac ggc
leu ile gln phe asp ala ala ile gln pro gly asp ser gly gly pro val val asn gly
721/241                               751/251
cta gga cag gtg gtc ggt atg aac acg gcc gcg tcc gat aac ttc cag ctg tcc cag ggt
leu gly gln val val gly met asn thr ala ala ser asp asn phe gln leu ser gln gly
781/261                               811/271
ggg cag gga ttc gcc att ccg atc ggg cag gcg atg gcg atc gcg ggc cag atc cga tcg
gly gln gly phe ala ile pro ile gly gln ala met ala ile ala gly gln ile arg ser
841/281                               871/291
ggt ggg ggg tca ccc acc gtt cat atc ggg cct acc gcc ttc ctc ggc ttg ggt gtt gtc
gly gly gly ser pro thr val his ile gly pro thr ala phe leu gly leu gly val val
901/301                               931/311
gac aac aac ggc aac ggc gca cga gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt
asp asn asn gly asn gly ala arg val gln arg val val gly ser ala pro ala ala ser
961/321                               991/331
ctc ggc atc tcc acc ggc gac gtg atc acc gcg gtc gac ggc gct ccg atc aac tcg gcc
leu gly ile ser thr gly asp val ile thr ala val asp gly ala pro ile asn ser ala
1021/341                              1051/351
acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc ggt gac gtc atc tcg gtg acc tgg
thr ala met ala asp ala leu asn gly his his pro gly asp val ile ser val thr trp
1081/361                              1111/371
caa acc aag tcg ggc ggc acg cgt aca ggg aac gtg aca ttg gcc gag gga ccc ccg gcc
gln thr lys ser gly gly thr arg thr gly asn val thr leu ala glu gly pro pro ala
1141/381
tga
OPA

```

SEQ ID N° 50F

FEUILLE DE REMPLACEMENT (REGLE 26)

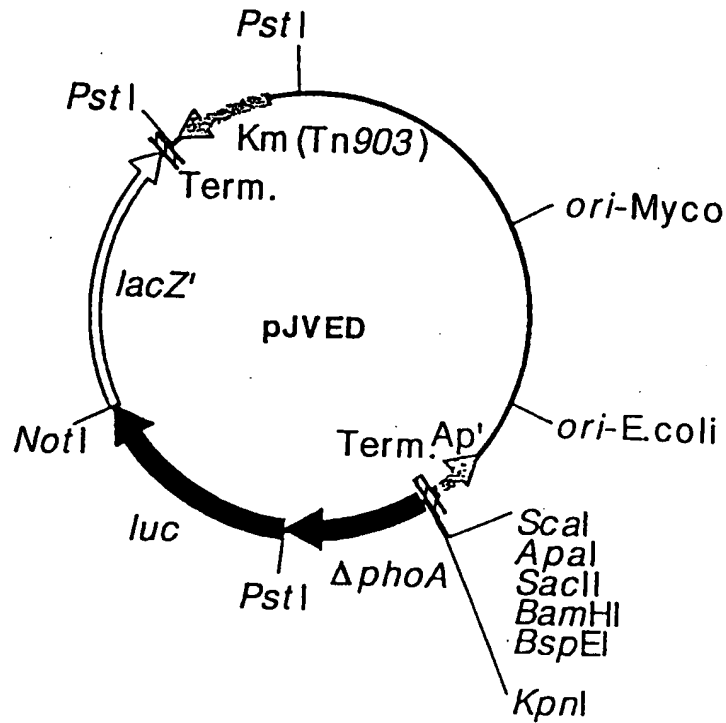


FIGURE 51A

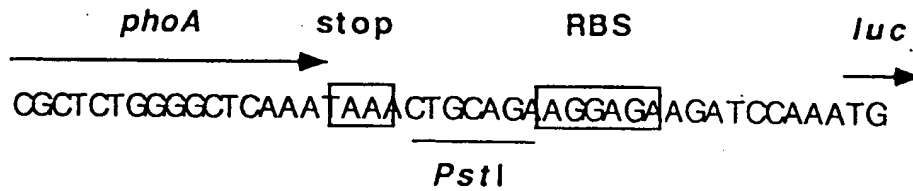
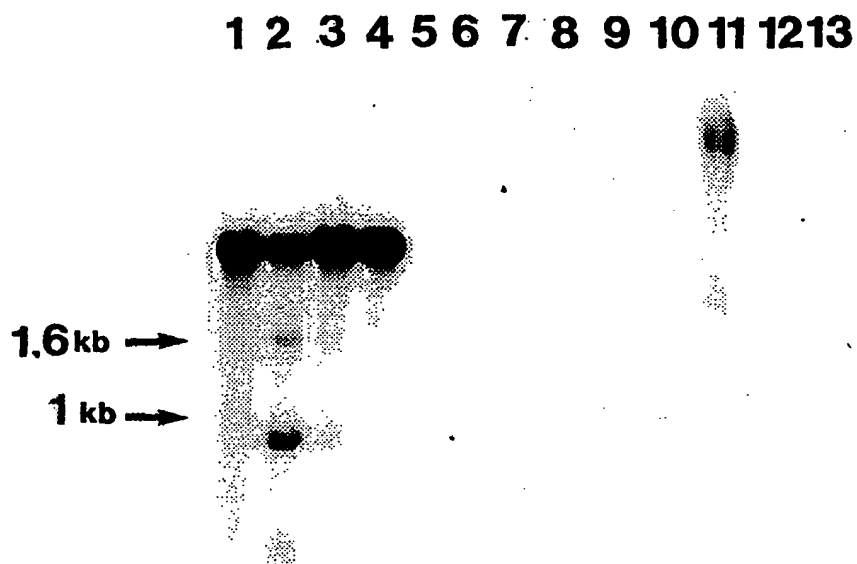


FIGURE 51B

182/185

Expériences d'hybridation moléculaire d'une sonde spécifique
du DP428 sur l'ADN génomique de différentes espèces de
mycobactéries



1: *M. tuberculosis* 2: *M. bovis* 3: BCG 4: *M. africanum* 5: cancelled 6: *M. fortuitum* 7: *M. simiae* 8: *M. avium* 9: *M. chelonae* 10: *M. flavescens* 11: *M. gordonae* 12: *M. marinum* 13: *M. kansasii*

FIGURE 52
FEUILLE DE REMPLACEMENT (REGLE 26)

183/185

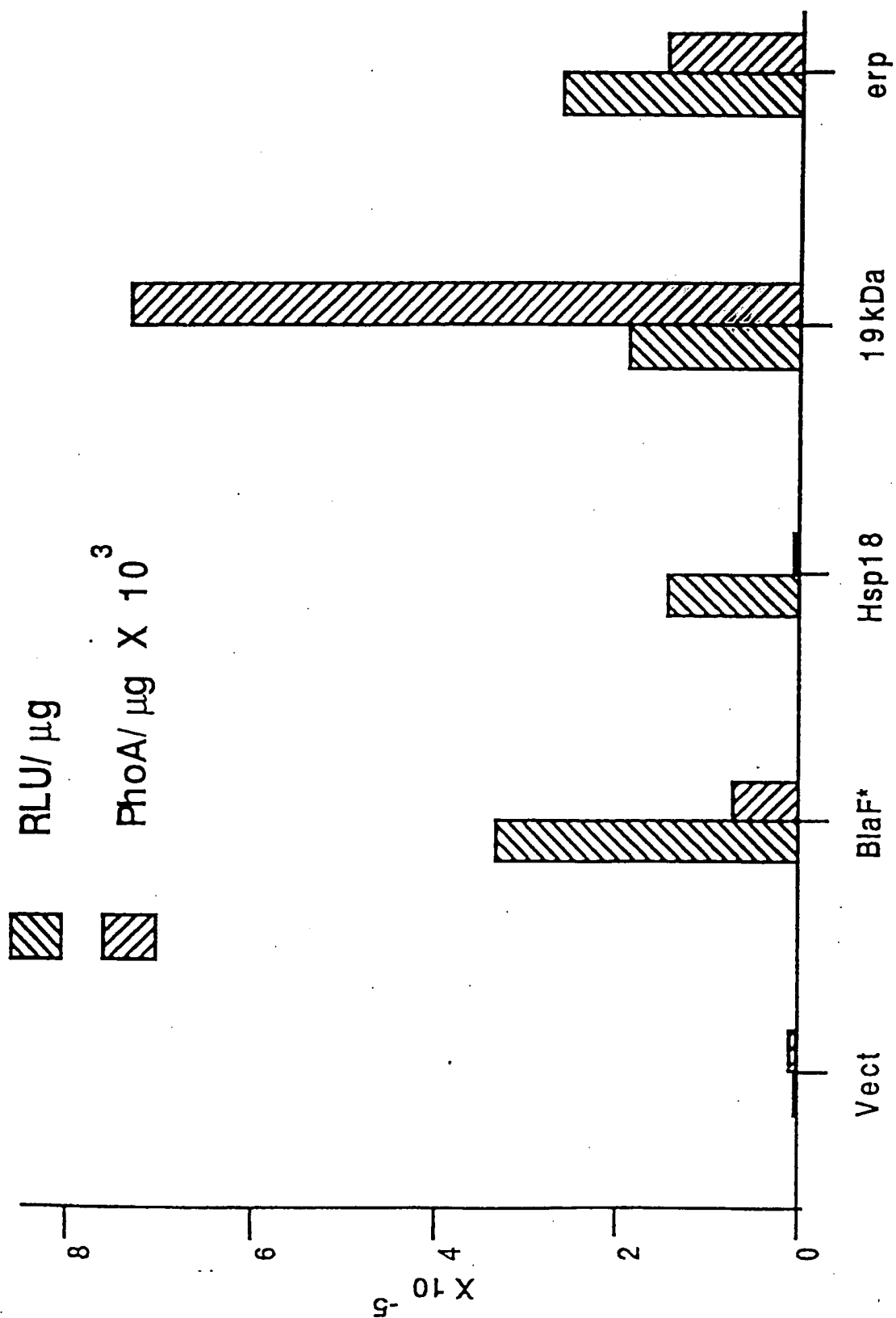


FIGURE 53

184/185

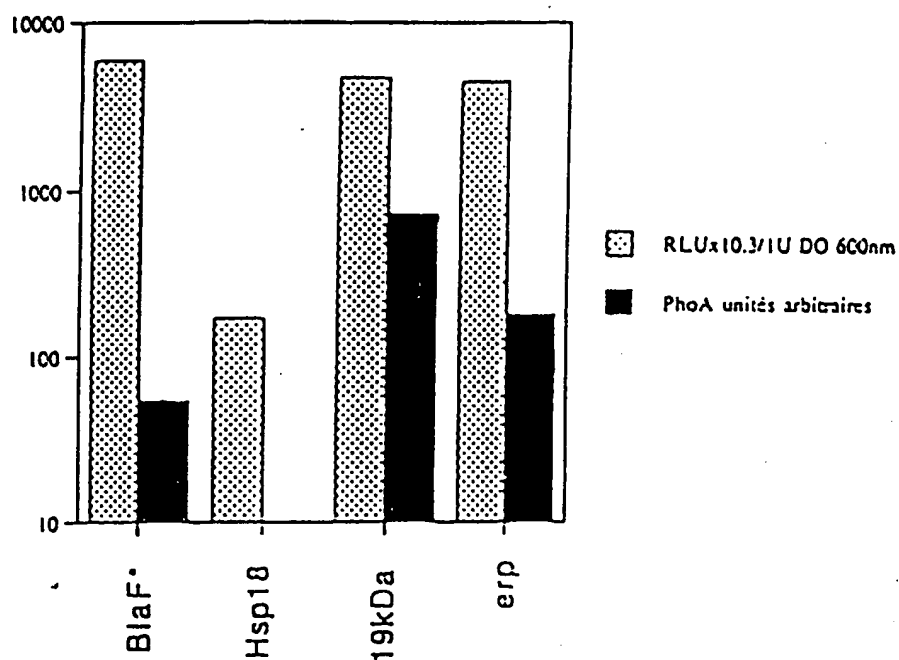
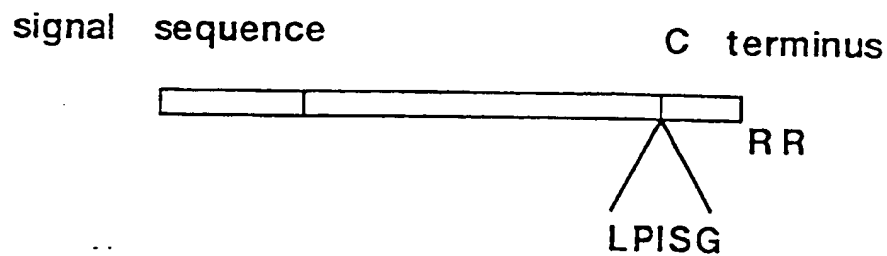
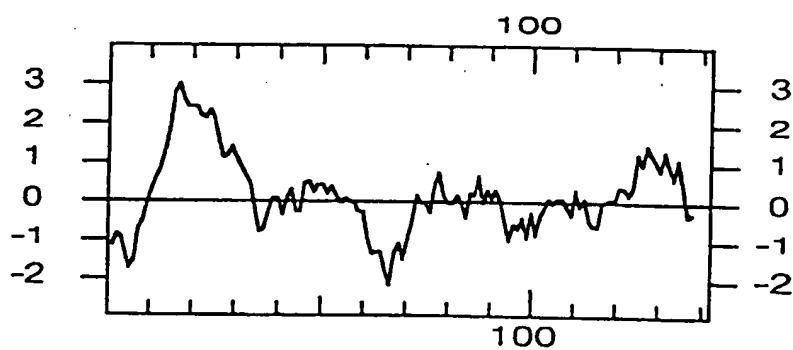


FIGURE 54



185/185

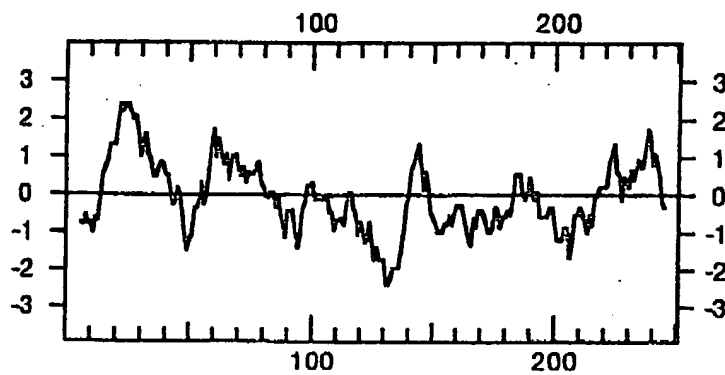


FIGURE 56

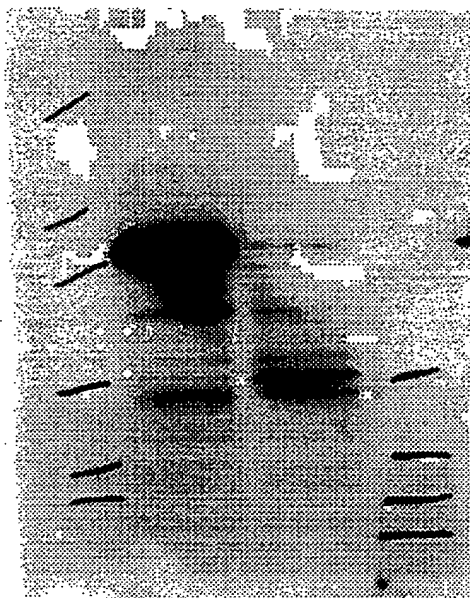


FIGURE 57A

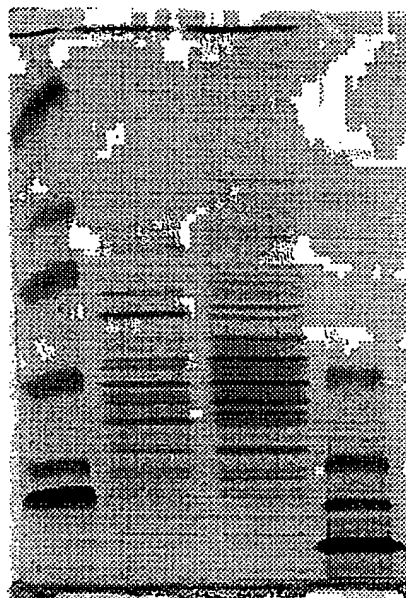


FIGURE 57B